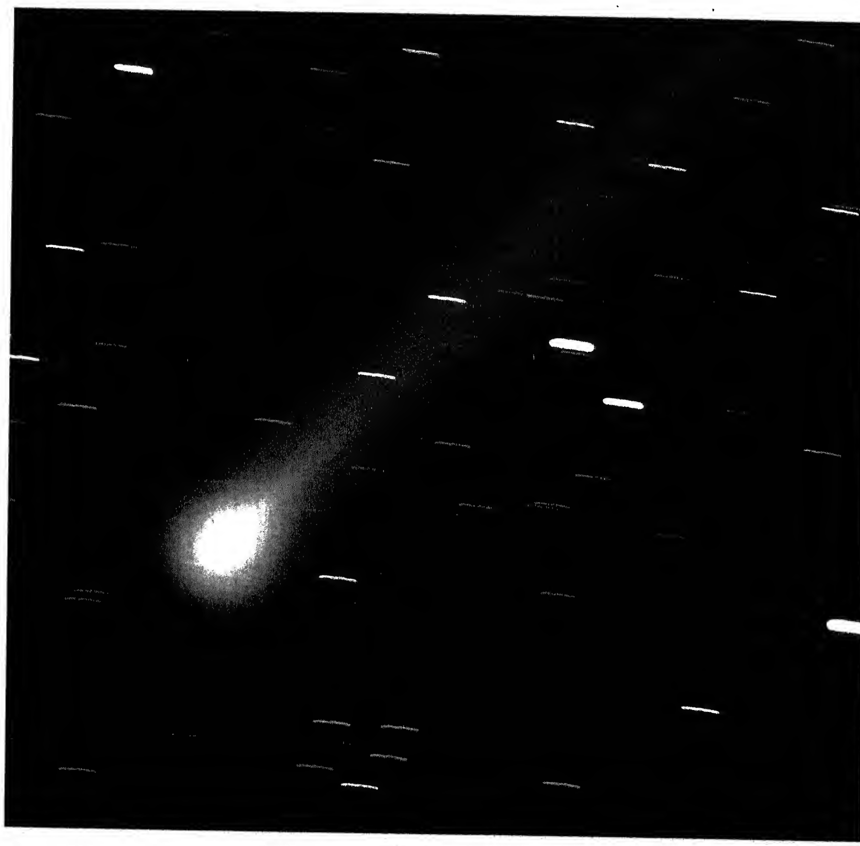
Resonance

June 1996

Volume 1 Number 6

journal of science education



Hybrid Cells and Human Genetics ❖ What Happened to Comet Hyakutake? ❖ Energy Storage and Retrieval ❖ Video-On-Demand ❖ The Geometry of Manifolds

Resonance is a monthly journal of science education published by Indian Academy of Sciences, Bangalore, India.

Editorial Office: Indian Academy of Sciences, C V Raman Avenue, PB No. 8005, Bangalore 560 080, India.
Tel:+91 (80) 3342310 / 3342546, Fax:+91 (80) 3346094, email:resonanc@ias.ernet.in

Editorial

N Mukunda, Chief Editor

With the appearance of this issue, ***Resonance*** completes six months of existence. At this time we express our sincere thanks and appreciation to Srinivas Bhogle who joined us when plans for ***Resonance*** were being made, and offered to work as Production Editor. He and his colleagues A S Rajasekar, V Ravi and R Guruprasad, all from National Aerospace Laboratories, Bangalore have brought skill and imagination to the design, layout and general appearance of ***Resonance***, and given it a truly distinctive quality. We editors have learned much from them in the process, and hope to maintain the standards of production set by them — may be after a while do even better, which should please them too. As they leave us, we make the transtition to a team wholly within the Academy. This may occasion some delays in the appearance of one or two issues, but hopefully no more.

Kapil Paranjape, in his series on Geometry, deals in this issue with the historic June 1854 inaugural lecture of Georg Friedrich Bernhard Riemann at Gottingen, titled "On the hypotheses which lie at the foundations of geometry." It is humbling to realise that Riemann gave this lecture, which so profoundly influenced the growth of both mathematics and physics for succeeding generations, so that he would then be permitted to teach at the University! His ideas were carried forward by Elwin Bruno Christoffel, and especially by the Italian geometers Gregorio Ricci - Curbaatro and Tullio Levi-Civita, culminating in their Absolute Differential Calculus. All this came in handy — "just what the Professor ordered" — when Albert Einstein was developing his general theory of relativity; though to be precise one must point out that Einstein's inspiration sprang from physical principles and ideas, and he was later introduced to the available mathematical tools by his friend Marcel Grossmann.

It is humbling to realise that Riemann gave his epic lecture on the foundations of geometry, which so profoundly influenced the growth of both mathematics and physics for succeeding generations, so that he would be permitted to teach at the University!

In this context it is difficult to avoid the temptation to recall a few other memorable lectures in the history of mathematics and science, each of which had profound consequences for later developments. One was the October 1872 lecture of Felix Klein at the University of Erlangen, titled "Comparative review of recent research in geometry", in which the characterization of each kind of geometry by its group of symmetries was established. This

Konrad Lorenz —
father to ethology and mother to ducks, geese, jackdaws, salamanders, fish and many more!

Konrad Zacharia Lorenz was born in 1903 as the son of a distinguished orthopaedic surgeon in Vienna. Like Darwin (and perhaps like so many of today's kids), he studied medicine in deference to the wishes of his father and like Darwin, his real love lay elsewhere - in this case, in the study of animal behaviour. But "study" is not how it began. Lorenz spent most of his life in the company of his favourite pets - ducks, geese, jackdaws, salamanders, fish and many more. Very early in childhood, when the scientific study of animal behaviour, or of anything else for that matter, was far from his mind, Lorenz and his childhood friend and future wife Gretel, took care of their pet ducklings by pretending to be mother ducks by living with them "a complete duck's life" and becoming "most thoroughly familiar with the whole repertoire of all the things a duck can say or do", without knowing that "this was to be called an 'ethogram' many years later".

It is now a matter of history that the science of the study of animal behaviour christened, *ethology*, was largely developed by Konrad Lorenz along with Niko Tinbergen and Karl von Frisch for which they won the Nobel prize in 1973.

Lorenz is best known for his discovery of *imprinting* the phenomenon by which young animals, especially birds, learn to recognize their mothers. Lorenz's discovery of course came from the fact that his ducklings and goslings became imprinted on him as he had replaced their mother early enough. Today imprinting is well understood as a form of learning possible only in early life and of enormous adaptive value.Young animals are not born with a mental image of their mothers but are born with an innate capacity to follow their mother or any moving object (including Lorenz) and take it to be their mother. In some species such "knowledge" gained during childhood, of what members of their species are supposed to look and sound like, is also used by the adult birds to correctly court conspecifics rather than some wrong species. This Lorenz discovered to his great delight when his hand reared jackdaws began to court him! Along with Niko Tinbergen, Lorenz laid the foundations of ethology particularly in relation to the organization of instinctive behaviour. Together they demonstrated the presence of motor programs, fixed action patterns and innate releasing mechanisms, phenomena that are still poorly

circle of ideas is often referred to as the "Erlangen Programme". Then in 1900 comes the famous Paris lecture by David Hilbert at the second International Congress of Mathematicians, titled "Mathematical Problems" and whose aim was no less than to list a set of problems (twenty three in all) which would engage the attention of mathematicians for the coming century. Sometime after the development of special relativity by Einstein in 1905, we

understood from a neurophysiological perspective.

Lorenz was a personality hard to ignore. His biographer Alec Nisbett describes him as "colourful, almost flamboyant; assertive in the manner of his speech and controversial in much that he says, but deeply interesting as an individual and as a man in relation to his concept of science." Lorenz's method of and motivation for doing science deserve our attention especially in today's fast-track, high-tech, publish or perish culture. Lorenz laments that "it is fashionable in science nowadays to experiment rather than to observe, to quantify rather than to describe ... descriptive science, based on plain unbiased observation is the very fundament of human knowledge ... I contend that not even a person endowed with the almost superhuman patience of a yogi could look at animals long enough to perceive the laws underlying their behaviour patterns. Only a person who looks with a gaze spellbound by that inexplicable pleasure we amateurs, we dilettante enjoy, is in a position [to do so]." For all his dislike of experiments, Lorenz admits "Niko [Tinbergen] and I were the perfect team. I am an ... amateur and prefer observing to experimenting ... Niko Tinbergen is the past master of the unobtrusive experiment ... We published in joint authorship, a paper that has become a classic."

But coming back to his motivation for doing science (notice that he calls himself an amateur), Lorenz's favourite quotation is that "Man is only then completely human when he is at play". He recalls his early childhood by saying that "As a very little boy, I loved owls and was quite determined to become an owl. In this choice of profession I was swayed by the consideration that an owl was not put to bed as early as I was". On retirement at the age of seventy one, Lorenz said " If you ask me what I have done throughout my life in the field of research and teaching, then I must honestly say: I have always done the things which at the moment I considered the greatest fun." I wonder if modern science, even ethology, any longer affords us the luxury of doing what we consider the greatest fun and yet make fundamental discoveries. Perhaps not, but the explanation that most of us find satisfying for this state of affairs is that our science is far too advanced to permit the kind of amateur play that Lorenz's infant ethology did. Indeed we often go so far as to believe that all branches of science are far too advanced to permit any place for Lorenz today. I hope that young readers of *Resonance* are *not* convinced!

Raghavendra Gadagkar

The purpose in recalling some of these historic lectures is to stimulate at least a few of our young readers to look for and go through the original texts; and also to convey a sense of wonder that profound and daring thoughts could be presented over an hour or so, which would then guide progress for years to come - a staggering realisation indeed.

have the September 1908 address by Herman Minkowski at Cologne on "Space and Time", containing the memorable line "Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

Skipping a couple of decades (and some other lectures), we remember next Niels Bohr's address, in September 1927 at the Alexandro Volta centenary meeting in Como, titled "The quantum postulate and the recent developments of atomic theory." It was here that he first presented his *complementarity principle*, as a general philosophical guide to the interpretation of quantum mechanics. It turned out though that not many in the audience could comprehend his ideas, and they had to be further elaborated later on. Then in August 1932 Bohr spoke to a gathering of physicians at Copenhagen on "Light and Life" — an attempt to apply the complementary principle to the understanding of life. Though his expectations have subsequently not been realised, this lecture deeply influenced one young listener — Max Delbruch — and induced him to turn from theoretical physics to biology, with well-known consequences.

The purpose in recalling some of these historic lectures is to stimulate at least a few of our young readers to look for and go through the original texts; and also to convey a sense of wonder that profound and daring thoughts could be presented over an hour or so, which would then guide progress for years to come — a staggering realisation indeed. Borrowing from Bernard Shaw: "The reasonable man adapts himself to the circumstances surrounding him. The unreasonable man adapts the circumstances to himself. All progress depends on the unreasonable man." So is it in science and mathematics, as in other endeavours.

> To explain all nature is too difficult a task for any one man or even for any one age. 'Tis much better to do a little with certainty, and leave the rest for others that come after you, than to explain all things.

> But the years of searching in the dark for a truth that one feels, but cannot express; the intense desire and the alternations of confidence and misgiving, until one breaks through to clarity and understanding, are only known to those who have themselves experienced it. *Albert Einstein (1933)*

Science Smiles

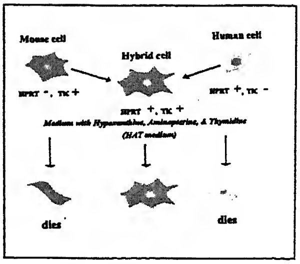
R K Laxman

The analysis does reveal that your boy has
a very high degree of criminal tendencies

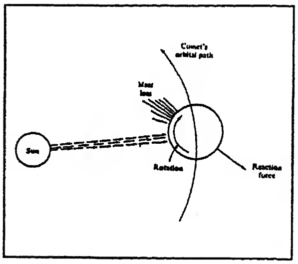
Resonance journal of science education

June 1996 Volume 1 Number 6

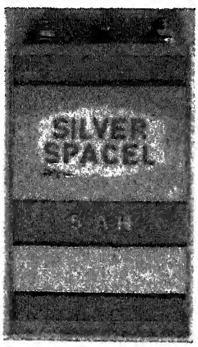
41

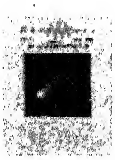


50



61





Front Cover
Comet C/1996 B2 (Hyakutake) on 23 March 1996 from Vainu Bappu Observatory, Kavalur, Indian Institute of Astrophysics (Photographed by K Kuppuswamy and Pavan Chakraborty; Courtesy : Indian Institute of Astrophysics, Bangalore 560 034.

Back Cover
Konrad Zacharia Lorenz (1903- 1986)

Origin (?) of the Universe

5. Observational Cosmology

Jayant V Narlikar

Jayant Narlikar, Director, Inter-University Centre for Astronomy and Astrophysics, works on action at a distance in physics, new theories of gravitation and new models of the universe. He has made strong efforts to promote teaching and research in astronomy in the universities and also writes extensively in English and Marathi for a wider audience on science and other topics.

This six-part series will cover: 1. Historical Background. 2. The Expanding Universe. 3. The Big Bang. 4. The First Three Minutes. 5. Observational Cosmology and 6. Present Challenges in Cosmology

In this part of the series we look at the present astronomical evidence both from distant parts of the universe as well as from our local region to test cosmological predictions. Included in this discussion are Hubble's law, source counts, angular diameters, the age of the universe, abundances of light nuclei and the microwave background. The significance of the findings for big bang models is discussed.

Cosmology As a Science

In the previous parts of this series we developed theoretical ideas in cosmology bringing the story so far to the standard big bang models. We saw that by using Hubble's law as the starting point in extragalactic astronomy we arrive at a picture of the expanding universe whose dynamical behaviour is modelled by Einstein's equations. The models of Alexander Friedmann developed back in 1922 still continue to serve as the basic models for cosmological studies.

We have been following the method of science in the study of the large scale structure of the universe. We began with a basic observation and then provided a theory to understand it. The scientific method then requires us to make new predictions to be tested by more sophisticated observations. This is what we will now proceed to do. We will follow the doctrine of Karl Popper who has argued that no scientific theory is ever proved. Its strength lies in making *disprovable* predictions. Thus we will consider the big bang cosmology like every scientific theory, to be on an indefinite probation. It continues to stay as a viable theory of the universe so long as the observational tests do not disprove it. By the same token, if we do encounter a conflict of the theory

with observations and the latter are seen to be correct, then we should be prepared to modify or abandon the theory.

In this part of the series we will discuss this confrontation between theory and observations. To this end we will consider two types of observations:

- *Type I :* These are based on surveys of distant parts of the universe. As light travels with a finite speed, these surveys tell us about the state of the universe in the remote past. Thus we can compare it with the present state and decide whether the changes seen are consistent with the predictions of the model.

- *Type II:* Here we need not look at distant parts of the universe; studies of the local environment can tell us whether it is consistent with the model.

We will follow the doctrine of Karl Popper who has argued that no scientific theory is ever proved. Its strength lies in making disprovable predictions.

Although several such checks and tests exist in the literature, we will only look at a few to see how cosmology is no longer merely speculative but is subject to the rigours of science. We will first consider some *type I tests.*

The Extension of Hubble's Law

The first measurement of importance in cosmology, as mentioned earlier, was Hubble's law. Hubble's original studies were confined to nearby galaxies and he got a value for the constant $H = 530$ kilometres per second per megaparsec. (The *parsec* is a unit of distance used by astronomers. Its value is approximately 3.10^{18}cm. Thus Hubble found that a galaxy at a distance of a million parsecs from us should have a radial velocity of recession of 530 km/second.

The questions that cosmologists would like to have answers to at this stage are:

(i) Is the value of Hubble's constant as given by Hubble in 1929 correct?

(ii) If we observe more remote galaxies would the velocity-distance relation still hold good?

Naturally, in the six and a half decades since Hubble's observations, astronomical techniques have improved enormously and one can carry out such observations on galaxies a thousand times farther away. Based on these measurements the answers to the two questions above are respectively *no* and *yes*. Let us see *why*.

• *Hubble's constant:* The most recent measurements of Hubble's constant have been carried out by the *Hubble Space Telescope* (HST) as one of its key projects. With its abilitv to observe objects at least fifty times fainter than the best ground based telescope, the HST is admirably suited for this programme. In 1994 the telescope was able to observe twelve Cepheid variable stars in the galaxy M 100 situated close to the centre of the Virgo cluster. The Cepheids exhibit a periodic variation in their luminosity. Moreover, there is a well established relation which tells us how the mean luminosity of a Cepheid can be determined from its period. So, by observing the period of a Cepheid, we know its intrinsic luminosity. By measuring its apparent brightness we can then determine its distance.

The most recent measurements of Hubble's constant have been carried out by the *Hubble Space Telescope* (HST). With its ability to observe objects at least fifty times fainter than the best ground based telescope, the HST is admirably suited for this programme.

Hubble himself had used the same method to measure distances of galaxies but these could not be very far away if the Cepheids in them were to be seen. Moreover, in many cases, he mistook other variable stars for Cepheids. And this led to an underestimate of distance and an overestimate of H. The HST determinations lead to an H value in the range 65-80 km/s/Mpc, that is, only about 12-15 per cent of the 1929 value!

It is as well to pause and review the difficulties in this basic measurement. The measurement of distance in astronomy has always been difficult, and the difficulty grows as we attempt measurements of more distant objects. The astronomer uses the 'standard candle' type argument; that is, if we have two similar sources of light, one bright and the other faint, the latter is

expected to be farther away than the former. The basis for this is the inverse square law of illumination *and the assumption* that both sources are intrinsically equally luminous. In practice, these premises may not hold; for example there may be absorption of light *en route* or the two sources may not be equally luminous.

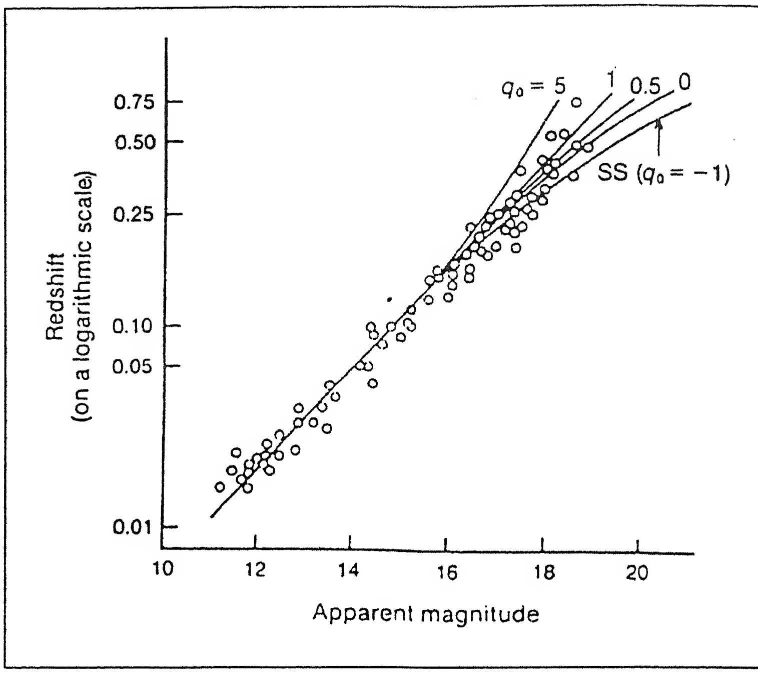
The second difficulty lies in the fact that in addition to the expansion of the universe the galaxies also possess 'peculiar' (i.e., random) motions within a cluster. When we measure the redshift of a galaxy and infer its radial speed, we are actually measuring the net speed. It becomes difficult to disentangle the true Hubble expansion speed from the net speed especially for relatively nearby galaxies. Thus our earth moves around the sun, which moves around the centre of the galaxy, which in turn is falling towards the centre of the Virgo cluster. (Recall that our cosmologies were based on the Weyl postulate which assumes that all such speeds are to be neglected.)

The measurement of distance in astronomy has always been difficult, and the difficulty grows as we attempt measurements of more distant objects.

Because of these difficulties the value of Hubble's constant has always remained controversial; even today another group of astronomers led by Allan Sandage believes that the true value of H is much lower, say in the range of 45-65 km/s/Mpc.Keeping in view the prevailing uncertainty, we will write H as 100h km/s/ Mpc, with h representing a fraction between 0.5 and 0.8.

- *Hubble's law at large redshifts:* Concerning the second question, observations of galaxies out to redshifts of the order unity have shown that a reasonably tight Hubble relation exists for them. As mentioned earlier, the astronomer does not measure distance directly but instead measures the faintness of a source. Allan Sandage has shown that if we pick the first ranked (i.e., amongst the brightest class) galaxies in clusters, they have approximately the same luminosity. That is, they serve as standard candles. Suppose a typical galaxy has luminosity L and is located at distance D from us. Then the flux from the galaxy will be $L/4\pi D^2$, as per the inverse square law of illumination. The astronomer measures the logarithm of this quantity on the so-called magnitude

Figure 1 The redshift magnitude relation for first ranked cluster member galaxies. A number of theoretical curves are superposed on the data. These curves are labelled by a parameter q_0 which measures the rate of slowing down of the expansion of the universe. It is called the **deceleration parameter** *and its value for the flat Friedmann model is 0.5. We will refer to the steady state model in the final part of this series. (Based on J Kristian, A Sandage & J A Westphal, The Extension of the Hubble Diagram-III.* **The Astro-physical Journal**, *221, 383 (1978).*



scale. Thus we would write the magnitude as

$$m = -2.5 \log (L/4\pi D^2) + \text{constant}.$$

The constant arises from normalization of the magnitude scale.

What is D? Recall that the Friedmann models use curved spaces. Indeed, Einstein's general relativity assumes that spacetime itself has curvature and that the geometrical properties of spacetime are determined by the amount of matter and radiation in the universe, through their pressure and density. Thus care is needed in computing D for each Friedmann model. This calculation, which we cannot go into here, not only uses the spatial properties but also the temporal ones that lead to the redshift. The result is that each model predicts a unique redshift-magnitude relation.

The relation found by Hubble agrees with the above relation for small values of D. As we go towards higher redshifts we expect to have different relations for different Friedmann models. Thus if we have the actual measurements of m and z for distant galaxies

we should be able to distinguish between the three types of models described in Part 3 of the series.

In the last few decades the m-z relation has been extended by Allan Sandage and others, to large values of these quantities, but several observational errors and uncertainties including those of distance determination intervene to prevent a decisive conclusion from being reached. The latest results suggest that models of type III are favoured by the data but considerable caution is still needed in basing our theoretical conclusions on such a result.

In the last few decades the m-z relation has been extended by Allan Sandage and others, to large values of these quantities, but several observational errors and uncertainties including those of distance determination intervene to prevent a decisive conclusion from being reached.

The Counts of Discrete Sources

Another way of determining which of the Friedmann models is correct is to test the volume-radius relationship of space. Thus, recalling Part 3 of the series, the type I Friedmann models use flat Euclidean geometry in which the volume V of a sphere of radius R is given by

$$V = (4\pi/3)R^3.$$

The relation is different for other models.

The astronomer can use this result to test the models in the following way. Suppose there is a population of discrete sources distributed uniformly in space. Then by counting such sources out to distance R, we get a number N proportional to V. If all sources are identical, then the flux S of radiation from the faintest source in this set would decrease with R, being proportional to R^{-2} in a Euclidean universe. Thus we expect a relationship of the form

$$N\,S^{1.5} = \text{constant}$$

for a Euclidean space. A more complicated relatlon ls derlved for each Friedmann model.

Hubble used this test in the thirties to select the correct Friedmann model by counting galaxies out to fainter and fainter magnitudes. In view of the relation between magnitude and the flux received

Hubble's objective of picking out the correct cosmological model still remains unattainable. This is because there are other uncertainties such as evolutionary changes in the source populations and the possibility of mistaking a nearby weak source for a distant powerful one.

from a light source encountered earlier, we see that in Euclidean space the number-magnitude relation will take the form

$$\log N = 0.6m + \text{constant.}$$

A departure from this relation could possibly indicate what cosmology we should have.

In practice this did not work, for two reasons. First, there were far too many galaxies to keep count. Second, one needs to go down to very faint magnitudes to notice any possible distinction between models. In Hubble's time these difficulties were insurmountable. But today, with the use of computers to count images of galaxies and sensitive detectors to record images of very faint objects, this test has again become feasible.

However, Hubble's objective of picking out the correct cosmological model still remains unattainable. This is because, there are other uncertainties such as evolutionary changes in the source populations and the possibility of mistaking a nearby weak source for a distant powerful one.

In the fifties, radioastronomers began to attempt this test for populations of radio sources. Radio sources are a much rarer species than galaxies and radio telescopes are capable of picking out distant radio sources more easily than their optical counterparts vis-à-vis galaxies. Thus it was felt that counts of radio sources may provide more decisive constraints on cosmological models. But here too, source counts have told us more about the physical properties of the source populations than about the large scale geometry of the universe.

Angular Diameters

In 1958, cosmologist Fred Hoyle proposed a new test of these models specially suited to radio sources. Its predictions are very striking and can be explained as follows.

It is a common experience here. on earth that if we view an object (like a tree or a house or a mountain) from increasingly larger

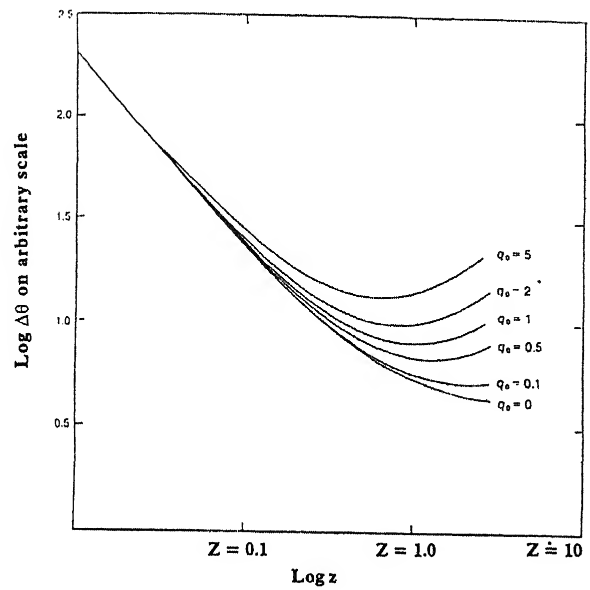


Figure 2 This figure illustrates how the angular diameter varies with redshift in a Friedmann model. Notice that at a specific redshift the angular diameter is the least in contrast to a Euclidean universe in which the angular diameter steadily decreases with distance. The label q_0 denotes the deceleration parameter as explained in the caption of the previous curve.

distances it appears smaller in size. This is because the angle θ subtended by it at our eyes gets smaller as our distance D from the object increases. If the linear size l of the object is small compared to D, and the angle is measured in radians, then a simple relation exists between θ and D in Euclidean geometry:

$$\theta = l/D.$$

This explains why distant objects look smaller since our brain relates the perceived size to the angle θ.

Hoyle showed that the corresponding situation in an expanding universe is more striking. Let us suppose we are viewing a population of identical objects located at varying distances in such a universe. Then we expect that their angular sizes should decrease with increasing redshift, since redshift is proportional to distance. Indeed, this is how θ behaves as z increases from zero. But the decline is not as fast as in the Euclidean case, and in most Friedmann models, the decline of θ is arrested at some redshift and thereafter θ increases with redshift! In the type I model ($k = 1$) discussed in Part 3 of the series, this minimum angular size is obtained at a redshift of 1.25.

It is not easy to measure the redshift of a radio source. It must first be optically identified and then its optical spectrum studied. Thus in many cases the redshifts are either not available or are estimated indirectly thus adding to the errors.

This nice and clean looking test also gets bogged down in uncertainty, because we cannot guarantee a sample of sources with nearly the same physical size, nor can we be sure that the sources we are looking at, at various redshifts, are not evolving. It is also not easy to measure the redshift of a radio source. It must first be optically identified and then its optical spectrum studied. Thus in many cases the redshifts are either not available or are estimated indirectly thus adding to the errors.

We will leave our discussion here, albeit in an unsatisfactory state because the tests proposed for distinguishing between cosmo–logical models are inconclusive. As our observations improve, we will be able to appreciate and allow for the various sources of errors and perhaps be able to draw some definitive conclusions.

We next briefly look at *type II tests* i.e. tests based on studies of our local neighbourhood. Here the idea is to find evidence that provides a consistency check on cosmological predictions.

Density of Matter in the Universe

Recall from our discussion of Part 3 that the three types of Friedmann models make three types of predictions about the present matter density in the universe. Let us define the *density parameter* Ω by the relation

$$\Omega = \rho / \rho_c$$

Thus, for open models $\Omega < 1$, for closed models $\Omega > 1$, while for the flat model, $\Omega = 1$. What is the 'true' value of Ω?

For this we first need to know the closure density. Using the formula in Part 3, with the Hubble's constant as 100/h km/s/Mpc, we get $\rho_c = 2.10^{-29} h^2 \mathrm{g\,cm^{-3}}$. Taking $h = 0.65$, the closure densitv of matter is approximately $8.5.\ 10^{-30} \mathrm{gcm^{-3}}$. So the question is, what is the actual density of matter in the universe?

This question demonstrates how a local measurement can be of cosmological significance. If we measure the density of luminous

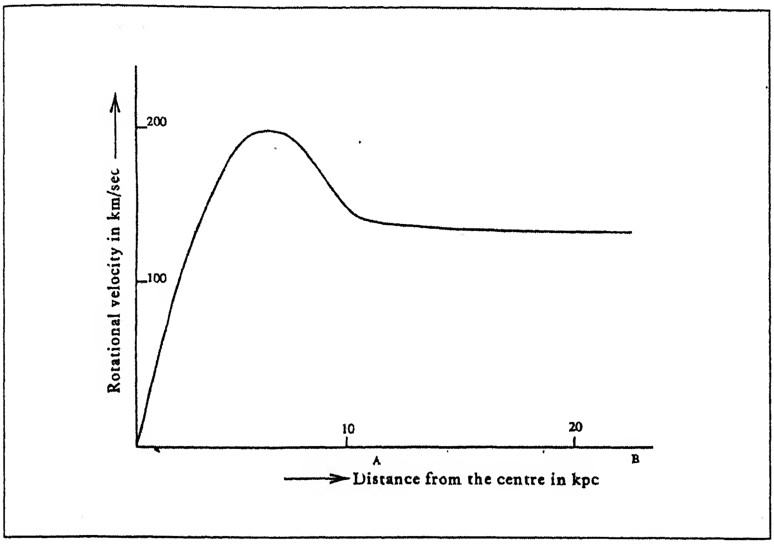


Figure 3 The typical flat rotation curve of a galaxy that indicates the presence of dark matter.

matter, we get a figure around 4. 10^{-31}g cm^{-3}, which is much less than the above closure density. However, this is only for matter seen as galaxies, clusters, intracluster medium, etc. There is also evidence for *dark matter.* Let us examine this briefly.

When we look at a galaxy we see a nebulous patch which is brightest in the centre and fades outwards. The luminosity of the patch arises due to stars which are most strongly concentrated at the centre and thin out towards the periphery. A typical spiral shaped galaxy may have stars distributed in a disc of radius, say, 15 kiloparsecs. Astronomers previously believed that the mass in a galaxy was confined to this visible disc. However, studies of clouds of neutral hydrogen emitting radiation at a wavelength of 21 cm showed their existence well beyond this radius. Such clouds are found rotating around the centre of the galaxy just as planets go around the sun. On the basis of Kepler's laws, we expect that if the attracting galactic mass is confined to this disc, the rotational speeds would decline as one examines cloud motions farther and farther out. *This is not the case!* The rotational speeds stay more or less constant for distances as far as 150 kpc.

The implication is, of course, that the 'driving mass'[1] extends well beyond the visible disc. Estimates show that it is not

[1] The phrase 'driving mass' refers to the total mass whose attraction on a particle keeps it in a circular orbit. To a good approximation, this is the mass enclosed in a sphere of the same radius.

Normally, on the basis of Kepler's laws we expect that the rotational speeds would decline as one examines cloud motions farther and farther out. *This is not the case!* The rotational speeds stay more or less constant for distances as far as 150 kpc.

negligible, but in fact exceeds the visible mass. Such unseen matter or dark matter seems to occur widely, being found not only around galaxies but also in the intracluster medium. The idea that dark matter exists in clusters of galaxies, however, dates back to Fritz Zwicky in the 1930s! Estimates vary as to how much of it there is. Conservatively we may say that dark matter may be ten times as abundant as luminous matter, thus raising our density estimates to around $4.\ 10^{-30}$g cm^{-3}.

We are still below the closure density, and so $\Omega < 1$, T but the gap now is not too wide and this has kept alive the other options $\Omega = 1$, or $\Omega > 1$. There are theoretical reasons for the first of these options which we will discuss in the final part of the series. We will also discuss the conjectures advanced on the composition of the dark matter.

Age of the Universe

All big bang models have finite ages which are not too difficult to compute; the age of a model is the time it has spent from the instant of big bang to the present epoch. The answer is best expressed in units of H^{-1}. The simplest answer is for the empty universe ($k = -1$, $\Omega = 0$ and is H^{-1}. For the flat universe, the answer is 2/3 of this value. In general, the age decreases as Ω increases.

Today, the most favoured Friedmann model is the flat one and for the value of Hubble's constant given by $h = 0.65$, the age is ten billion years. Although this value comfortably exceeds the age of the earth (about 4.6 billion years), it is not high enough to accommodate old stars. For example, age estimates of stars in the globular clusters which are found to have evolved to the stage when their hydrogen fuel is finished, are as high as 12–18 billion years. Thus only those models which have a very low value of Ω might possibly survive.

This has been a serious difficulty for all the standard models

discussed so far and we will return to this issue in the final part of this series.

Abundances of Light Nuclei

A basic problem of cosmology is to understand how matter came into existence and how it acquired its present elemental composition. We saw in Part 4 that two different processes may have been at work, one acting in stars and the other in the early universe. The latter is believed to be responsible for the production of light nuclei, mainly helium, deuterium and a few others.

As far as helium is concerned, its abundance by mass from primordial nucleo synthesis is expected to be in the range of 23-24 per cent, increasing slightly as the baryonic density (that is, the density in the form of baryonic particles like protons and neutrons) of the universe increases. A strong point in favour of the big bang picture is that measurements of helium abundance from various spectroscopic data are in broad agreement with this range. Moreover, the theoretical value is sensitive to the number of neutrino species that existed in the early universe and is based on the assumption that there are three species of neutrinos. The calculation described in Part 4 would yield a somewhat higher value of helium abundance (0.25-0.26, instead of 0.23-0.24) than if there were, say, four types of neutrinos. Thus one could claim that the primordial nucleosynthesis calculation predicts that there should be three neutrino species, a result borne out by particle accelerator experiments.

A basic problem of cosmology is to understand how matter came into existence and how it acquired its present elemental composition; two different processes may have been at work, one acting in stars and the other in the early universe.

Coming to deuterium, the result is very sensitive to the baryonic density of the universe. The observed abundance of deuterium is very small, being in the range $9.\ 10^{-6}$ to $3.5.\ 10^{-5}$. The abundance predicted bv the standard models declines with increasing baryonic density and matches this range provided the density does not exceed $7.\ 10^{-31}$gcm^{-3}. Beyond this density any deuterium produced is quickly destroyed. So, we have a problem. Evidence for dark matter suggests a density higher than this limit. So we

have to think of this matter in a non-baryonic form. This 'non-baryonic' matter would not react with deuterium in the early universe. So measurements of deuterium do not constrain the quantity of dark matter present.

There are some additional constraints from observations of other light nuclei but they are more subtle and we will not discuss them here.

Microwave Background

As mentioned in Part 4 of this series, the discovery of the microwave background is the best evidence for big bang cosmology. It has been shown to have a Planckian (black body) spectrum with a temperature of 2735 K, thanks to the extensive work of the COBE satellite. Nevertheless, the following information emerges from measurements of this radiation background.

The radiation background would have been completely isotropic if we as observers were at rest in the cosmological frame (i.e., had no random motion relative to the Hubble motion of expansion). However, we do have motion and so the radiation background should show a 'dipole anisotropy' in our frame. Such an anisotropy has indeed been found with temperature being direction dependent:

$$T(\theta) = T_0 + T_1 \cos\theta.$$

Suggested Reading

J V Narlikar. The Structure of the Universe. Oxford. 1977.

David Layzer. Constructing the Universe. Scientific American Library. 1984.

JV Narlikar. Introduction to Cosmology. Cambridge. 1993.

From all these observations hopefully a final picture would emerge.

Address for correspondence
Jayant V Narlikar
IUCAA. PB No.4,
Ganeshkhind.
Pune 411 007, India

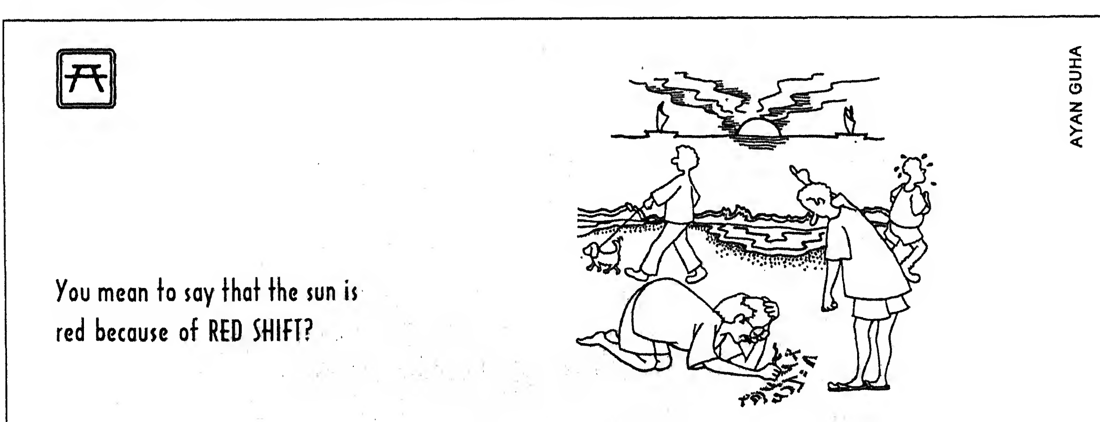
Here T_0 is the average temperature which is constant in all directions and $T_1 \cos\theta$ is the variable part, being maximum in the direciion $\theta = 0$, and minimum in the opposite direction. T_1 is approximately $3.\ 10^{-3}$. *The direction in which* $T_1 \cos\theta$ *is maximum is the one in which the earth is moving.*

The second type of anisotropy is on a small angular scale of about ten degrees which shows a fluctuation of $\Delta T / T$ of the order of $6.\ 10^{-6}$. This was the second major discovery of COBE which was announced in 1992, and it held out the hope of linking this fluctuation to the theories of structure formation. We will return to this in the final part of the series.

Concluding Remarks

As far as type I tests are concerned, because of various uncertainties, studies of distant parts of the universe have so far failed to tell us which type of Friedmann model (if any) we live in. The type II tests are somewhat more clearcut, but their verdict has been mixed. On the one hand, we encounter a serious 'age' problem, whereas on the other, the observations of microwave background and light element abundances are in favour of the hot big bang cosmology.

In the final part of the series we will consider the overall situation in cosmology today and address some fundamental issues.



Algorithms

3. Procedures and Recursion

R K Shyamasundar

R K Shyamasundar is Professor of Computer Science at TIFR, Bombay and has done extensive research in various foundation areas of computer science.

In this article we introduce *procedural abstraction* and illustrate its uses. Further, we illustrate the notion of *recursion* which is one of the most useful features of procedural abstraction.

Procedures

Let us consider a variation of the problem of summing the first M natural numbers. The problem is: *Compute the partial sums of all numbers from 1 to M.*

That is, we have to compute

$$\Sigma_{i=1}^{j} \; i \qquad (1)$$

for all j such that $1 \leq j \leq M$. This means we have to compute M sums according to equation (1).

A Naive Solution. Let us recall the program (discussed in *Resonance,* Vol.1, No.3) for summing N numbers. The code segment for summing N numbers is given in *Table 1.* Using the segment "code", we can obtain the algorithm for the problem on

Table 1. Program segment "code"

```
count: = 0;
sum: = 0;
i: = 1;
while (count < N) do
sum: = sum + i;        (* sum contains the sum of first i numbers *)
i: = i + 1;            (* increment i to get the next number *)
count: = count + 1;    (* count keeps count of the numbers added *)
endwhile;              (* sum contains the sum of first N numbers *)
print sum;
```

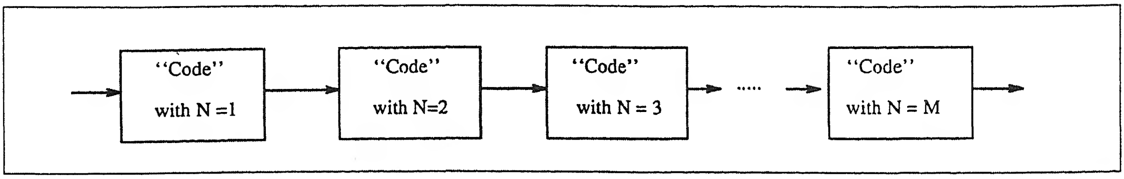


Figure 1 A naive solution for summing numbers.

hand as follows:

Place M copies of the program segment "code"(*Table 1*) in the form shown in *Figure 1*. Here, "code" with $N = j$ corresponds to the code shown in *Table 1* with N replaced by j.

In the program shown in *Figure 1*, we have repeated the algorithm M times and we can make the following observations. Each block is essentially a different instance of "code"; that is, the objects differ by the value to which N is initialized before the execution of the "code" block. Thus, we can now avoid the repetition of the "code" blocks by using an additional loop for setting the values of N. The program corresponding to such a refinement is shown in *Table 2*.

It uses an outer loop on the code segment "code" for passing the appropriate value for N. It is convenient to give the "code" a name so that it can be referred to (or used) later. In other words, the

Table 2. Program for finding all the partial sums

```
        N:=1;
while N ≤ M do
            count: = 0;
            sum: = 0;
            i: = 1:
          while (count <N) do
            sum: = sum + i;        (* sum contains the sum of first i numbers *)
            i: = i + 1;            (* increment i to get the next number *)
            count: = count + 1;    (* count keeps count of the numbers added *)
          endwhile:                (* sum contains the sum of first N numbers *)
        print sum;
        N:=N+1;
endwhile
```

Table 3. Procedural abstraction for summing numbers

```
procedure SUMMING (N: integer);
    count: = 0;
    sum: = 0:
    i: = 1;
    while (count < N ) do
        sum: = sum + i;
        i: = i + 1;
        count: = count + 1;
    endwhile:
    print sum;
endprocedure
```

Procedural abstraction not only makes the program concise but also easily comprehensible; the latter aspect is very important for verifying the correctness (either formally or informally) of the program.

name given refers to the object obtained by abstracting away the differences among the different instantiations of the repeated code; this is referred to as *procedural abstraction* and the name given is referred to as the *procedure name*. To account for the differences among the distinct instantiations, we list the variables of the program (i.e., the aspects in which the several objects differ) along with the name of the procedure. The variables thus listed along with the procedure are referred to as *formal parameters*. These variables/parameters can further be divided into *input* and *output* variables; the *input* parameters are those through which you provide the input values to the program. These values form the basis on which the outputs are to be computed. In other words, procedures could be treated as relations between the input and the output parameters. Such an algorithmic segment is called a *procedure* (or subroutine) and the variables of the algorithmic segment are referred to as formal parameters. The procedural form is shown in *Table* 3. The body (i.e., the code between the keywords procedure and endprocedure of the procedure) is the same as the "code" given in *Table 1*.

The identifier "SUMMING" is the name given to the algorithmic segment and is also referred to as the name of the procedure. In the procedure SUMMING, N denotes a formal variable/parameter

Table 4. Computing N sums using procedure SUMMING

```
N: = 1;
while N < M do
    call SUMMING ( j ).
    N: = N+1
endwhile
```

which can differ from one invocation to another. The notation *N: integer* indicates that N can take any value from the domain of integers. In other words, *Table 3* defines procedure SUMMING. Having defined a procedure, we can use it as if it were another *basic command*. To clearly distinguish it from the basic commands, we use the keyword "call" to indicate its usage. For example, "call SUMMING(100)", corresponds to executing the above procedure with the initial input value of N equal to 100. Now using the procedural form, the program for computing (1) (that is, computing all the intermediate sums up to M) can be written as shown in *Table 4*.

Procedural abstraction leads to saving space for storing program code, easy comprehension and a good structure.

In *Table* 4, the command "call SUMMING (j)" denotes the execution of the procedure SUMMING with variable (i.e., the formal parameter) N taking the value of j for each call; in this command j is referred to as the *actual parameter* as it is this value that is used in the execution of the procedure SUMMING. The keyword "call" denotes the invocation of the procedural segment. This keyword is omitted in several representations since it can be understood implicitly. It can be easily seen that the algorithms shown in *Tables* 2 and 4 are concise. Assuming one has understood "code", it can be said that the program shown in *Table 4* is more comprehensible than the one shown in *Table 2*. Thus, procedural abstraction not only makes the program concise but also easily comprehensible; the latter aspect is very important for verifying the correctness (either formally or informally) of the program. To summarize, procedural abstraction is based on the two principles indicated on the next page:

- Concentrate on the properties shared by several objects by abstracting away the differences.
- A succinct parametrization of the differences.

Procedural abstraction leads to saving space for storing program code, easy comprehension and a good structure. In fact, it becomes very handy when the same code-segment can be invoked for different purposes; these aspects as well as the power of parameters will become clear in the next article when we discuss a simple logo like programming language.

One of the usual techniques of problem solving is to break the problem into smaller problems. From the solution of these smaller problems one obtains a solution for the original problem.

Recursion

One of the usual techniques of problem solving is to break the problem into smaller problems. From the solution of these smaller problems, one obtains a solution for the original problem. Consider the procedural abstraction described above. It is possible to visualize the given procedure as being decomposed into a set of procedures. It may so happen that a smaller procedure (i.e., the sub-problem) is also of the same form as the original procedure, except that in '*measure*' it is '*smaller*' than the original. Assuming that we know the solution of problems for a certain finite set of *base cases,* we can then obtain a clear solution for the original problem or the procedure. A procedural abstraction which refers to itself is called *a recursive procedure.* We illustrate this powerful concept through the following example.

Example 1 (Towers of Hanoi) : This example is based on an ancient puzzle originating in a monastery in Tibet. We are given three

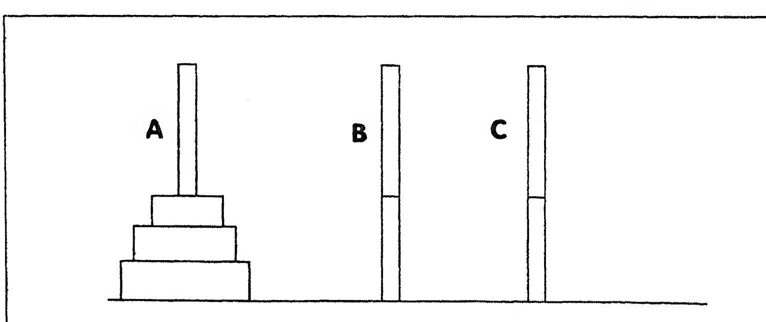


Figure 2 Towers of Hanoi problem.

rods and n disks of different sizes. The disks can be stacked on the rods, thereby forming 'towers'. Let the N disks initially be placed on rod A in the order of decreasing size as shown in *Figure 2*. The task is to move the N disks from rod A to rod C such that they are ordered in the original way. This has to be achieved under the following constraints:

- In each step only the topmost disk can be moved from one rod and placed on top of the disks on another rod.
- A disk may never be placed on top of a smaller disk.
- Rod B may be used as an auxiliary store.

The problem is to find an algorithm which performs this task. First let us consider the solution for $N = 2$. The solution is trivial : shift the smaller disk on A to rod B; then shift the larger disk on A to rod C; now shift the smaller disk on B to rod C (on top of the larger disk). Let us abstract it by the procedure *move_rod* (N_0 *A, B, C)* where N_0 (equal to 2 in this case) is the number of disks on A; B and C do not have any disks initially; and finally all the disks are transferred to C and they are in proper order. Let us see how we can arrive at the solution for other values of N using mathematical induction.

A procedural abstraction which refers to itself is called a *recursive procedure.*

Now we have a solution for the base case: *move_rod* (N, A, B, C) is certainly possible for $N = 2$ (base case). For inductively arriving at the successive steps, we have to derive the solution for $N_0 + 1$ from the solution of N_0.

Steps for $N_0 + 1$: The computation of *move_rod* ($N_0 + 1, A, B, C$) can be derived from *move_disk* ($N_0 + 1$,A, B, C) by the following steps:

- *move_rod* (N_0, A, C, B);
  N_0 disks are moved from A to B using C as auxiliary rod.
- *move_disk* (A, C);
  $(N_0 + 1)$th disk is moved from A to C directly.
- *move_rod* (N_0,B, A,C)
  The N_0 disks which are in proper order are transferred from B to C using A as an auxiliary rod.

Table 5. Procedure for the Towers of Hanoi problem

Steps:

```
procedure move_rod (N:Nat_Number A:name, B:name, C:name);
  if N > 1 then
      move_rod (N - 1, A, C, B);
      move_disk (A, C)
      move_rod (N - 1, B, A, C)
  else move_disk (A, C)
endprocedure
```

It may be noted that *move_disk* can be treated as a basic command.

The Towers of Hanoi problem is based on an ancient puzzle originating in a monastery in Tibet.

What is special in these steps? We have split the original problem into problems of smaller size. Further, the solution of the smaller problem is obtained by invoking the same procedure with appro–priate input. That is, *the main procedure calls itself.* This aspect of the procedures where one uses the ability of a procedure to *call itself* is referred to as *recursion.* This is one of the very important features of programming. The algorithm is shown in *Table 5.*

Trace of the steps for N = 4

- Step 1:
  move_rod (3, A, B, C)
- Step 2 : The above call leads to ($N>1$ holds):
  move_rod (2, A, C, B)
  move_disk (A, C)
  move_rod (2, B, A, C)
- Step 3 : The call to *move_rod (2, A, C, B)* leads to (N> 1 holds) :
  move rod (1, A, B, C)
  move disk (A, B)

move_rod (1, *C*, *A*, *B*)
move_disk (*A*, *C*)
move_rod (2, *B*, *A*, *C*)

- Step 4 : The call to *move_rod* (1, *A*, *C*, *B*) leads to ($N=1$ holds):

  move_disk (*A*, *C*)
  move_disk (*A*, *B*)
  move_rod (1, *C*, *A*, *B*)
  move_disk (*A*, *C*)
  move_rod (2, *B*, *A*, *C*)

- Step 5 : Taking the basic actions on the first two calls to *move_disk* leads to:
  {The smallest disk has been moved to peg C (corresponding to *move_disk* (*A*, *C*,)};
  {The next smallest disk (as the first has already been moved) has been moved to peg B (corresponding to *move_disk* (*A*, *B*))}

  move_rod (1, C, *A*, *B*)
  move_disk (*A*, *C*)
  move_rod (2, *B*, *A*, *C*)

At the end of step 5 the smallest disk is in peg C, the next larger in B and the largest disk remains at peg A.

- Step 6 : The call to *move rod* (1, *C*, *A*, *B*) ($N=1$ holds) leads to : {The smallest disk has been moved to peg C and the next smallest disk is on peg B and hence, the largest (3rd) is on peg A};

  move_disk (*C*, *B*)
  move_disk (*A*, *C*)
  move_rod (2, *B*, *A*, *C*)

- Step 7 : Executing the basic action *(move_disk)* leads to:
  {The smallest disk has been moved to peg C and the next smallest disk is on peg B and hence, the largest (3rd) is on peg A};

{The smallest disk on peg C is moved to peg B on top of the disk larger than it (corresponding to *move_disk (C, B))* };
{The largest disk on peg A is moved to peg C (which is empty now) corresponding to *move_disk (A, C)* };

move_rod (2, *B*, *A*, *C*)

- Step 8 : With the largest disk on peg C and the other disks on peg B (in the appropriate order), a call to *move_rod* (2, B, A, C) leads to:

  move_rod (1, *B*, *C*, *A*)
  move_disk (*B*, *C*)
  move_rod (1, *A*, *B*, *C*)

- Step 9 : With the largest disk on peg C and the other disks on peg B (in the appropriate order), (and hence, peg A is empty) the three basic actions can be rewritten as follows:
  {move the disk on B (the smallest) to peg A (corresponding to *move_disk (B, A))*};
  {move the disk from B (the second largest) on to the top of C already containing the largest disk (corresponding to *move_disk (B, C))*};
  {move the smallest on peg A onto the top of peg C already containing the other two disks in appropriate order (corresponding to the call *move_disk (A, C))*};

At the end of step 7 peg A is empty. Peg C has the largest disk and peg B the two disks in the right order.

Now, we have realized the objective of transferring the disks from peg A to C as per the protocol. Further, the program terminates as there are no calls left.

Now let us see what happens when the number of disks is less than 3. The case when $N = 1$ is trivial, as the disk is transferred straight from peg A to peg C. Now, let us consider the case $N = 2$. From the procedure, it can be seen that the call *move_rod* (2, *A*,*B*,*C*) reduces to

move_rod (1, *A*, *C*, *B*);
move_disk(*A*, *C*);
move_rod (1, *B*, *A*, *C*);

This further rewrites into:

mbve_disk (A, B);
move_disk (A, C);
move_disk (B, C);

The basic actions lead to:

move the smallest disk from A to B (corresponds to *move_disk (A, B));*
move the largest disk from A to C (corresponds to *move_disk(A, C));*
move the smallest disk from peg B to peg C on top of the largest disk (corresponds to *move_disk (B,C)).*

Example 2. *Recursive program for gcd*

Let us see from the analysis done earlier, whether we can arrive at a recursive program for computing the gcd of two positive numbers. From the earlier discussion, we have:

$gcd(m,n) = gcd\ (m \text{ rem } n, n).$

A recursive algorithm to compute gcd is very elegant.

Now, we can derive a recursive program from the following observations :

- To simplify, let us replace m rem n by $m - n$; this step should be convincing since division could be treated as repeated subtraction.

Table 6. A recursive gcd program

```
procedure gcd (m:integer, n:integer);
  if m=n then gcd is n
  else if m > n then gcd(m-n,n)
       else gcd(m, n-m)
       endif
  endif
endprocedure
```

- In the algorithm discussed in the second article (*Resonance*, Vol.1, No.3, 1996), the gcd algorithm terminates when the remainder becomes zero. Since we are using subtraction, it can be easily seen that the condition can be replaced by $m = n$.
- We should subtract the smaller number from the larger number. Thus, the roles of m and n may have to be reversed. Fortunately, $gcd(m,n) = gcd(n,m)$.

The program is shown in *Table 6*.

Discussion

In the previous sections, we have illustrated the advantages of procedural abstraction and introduced recursive procedures. The trace of the various invocations of the procedure calls, for the Towers of Hanoi example, shows how the procedures are invoked with new parameters. In a sense, one can consider the code yet to be executed as a *push-down stack* of procedure calls to be executed; in a push-down stack, you can access only the topmost element and hence we will be executing a procedure which entered the stack last (more about such data-structures will become clear in the forthcoming articles). Thus, one can assume that the program has terminated once the stack is empty and the last procedure has terminated. It can be observed that the recursive program for gcd looks simple and easy to understand. However, from this observation, we should not conclude that whenever possible one should use a recursive program. These aspects will become clear from the subsequent articles in this series.

Address for correspondence
R K Shyamasundar
Computer Science Group
Tata Institute of
Fundamental Research,
Homi Bhabha Road
Mumbai 400 005, India.

Suggested Reading

E W Dijkstra. A Short Introduction to the Art of Programming. CSI Publication. 1977.

R G Dromey. How to Solve it by Computer. Prentice Hall International. 1982.

D Harel. Algorithmics: The Spirit of Computing. Addison-Wesley Publishing Co. Inc. 1987.

This is one ofthe most lucidly written books on the topic.

Geometry

5. Enter Bernhard Riemann

Kapil H Paranjape

In the previous article the author examined curves and surfaces. One might hope to continue by analogy in many dimensions. The concept of working in many dimensions is so bewildering (yet today so matter-of-course) that it needed the genius of Bernhard Riemann to show us exactly how it can be done. In just one lecture on the foundations of geometry he completely changed our way of thinking. Later geometers were to spend entire lifetimes trying to finish what Riemann had begun. Some even see the genesis of *General Relativity* in his lecture.

After spending about a decade at the School of Mathematics, TIFR, Bombay, Kapil H Paranjape is now with Indian Statistical Institute, Bangalore.

A Manifold is a "Curvey-Curvey" Thing

Manifold: An object that displays extendibility in many dimensions, i.e. a multidimensional object.

How does one think about a higher dimensional object? This is the first question that Riemann addresses in his great lecture "On the hypotheses that lie at the foundations of geometry"[1]. While there are a number of objects that we can examine physically in dimensions up to and including three it is difficult to think of dimensions beyond *except abstractly* —in one's mind. Riemann suggests the following approach as one possible method. We can think of a curve as that which is traced out by a moving (or changing) point, a surface as that traced out by a changing curve, and so on. A manifold of dimension n is traced out by a changing manifold of dimension n–1. As we study such manifolds and learn their properties we will 'see' them more clearly.[2]

Analytically these n dimensions can be thought of as repre-

[1] This lecture was made particularly difficult to understand because Riemann had to lecture to a general (non-mathematical) audience and so he had to avoid too many symbols. This unfortunate circumstance continues to this day!

[2] This is a common approach in mathematics — we take a plausible definition and then work out all the properties. This gives us a feel for what the definition really entails.

[3] For more knowledgeable readers — we are following the convention of Riemann's lecture in that the study is only 'local' at the moment. Moreover I beg forgiveness of such readers for entirely neglecting analytical difficulties like the type of functions involved — smooth, analytic etc.

sented[3] by functions $x_1, \ldots, x_n$ on the manifold called the 'coordinate' functions. Specifying the values of each of these n functions simultaneously, uniquely specifies a point within the manifold. Thus other functions on the manifold can be thought of as functions $f(x_1, \ldots, x_n)$ of n variables. In fact functions of many variables had been studied before Riemann—it was his *geometric* interpretation that gave these earlier results more meaning and scope, and thus led to a spurt in the study of such functions. One point that needs to be emphasized, is that there is no obvious choice of coordinates—indeed there could be a number of such choices as we see even in the case of dimensions 2 and 3. We will try to examine properties that do not depend on any particular choice of coordinates.

The above definition gives us the property called *extension* by Riemann; it says that around any point our manifold extends in n independent directions. But he points out that this is not enough to specify the geometry of the manifold. The additional *metric* property is one that allows us to measure distances. In a radical departure from earlier work (except perhaps that of Gauss) he specifies distance by assigning magnitudes to 'line elements', which we have called velocity vectors. In other words, if one could measure the speed $s(t)$ at every instant t as we travel along a curve, then we could compute the distance we have travelled as $\int s(t)\, dt$. The distance between two points is then the *infimum* (or greatest lower bound) of the distances travelled along various curves between two points. From Gauss' theory of surfaces the length of a line element $dx = (dx_1, \ldots, dx_n)$ is specified by a form[4] $(\sum g_{ij}(x)\, dx_i\, dx_j)^{1/2}$. This means that the length of a curve $(x_1(t), \ldots, x_n(t))$ from t_0 to t_1 is given by the integral

[4] Riemann actually does consider more general forms briefly but we will skip them here for the sake of brevity.

$$l = \int_{t_0}^{t_1} \sqrt{\sum_{i,j=1}^{n} g_{ij}(x(t))\, \dot{x}_i(t)\, \dot{x}_j(t)}\; dt$$

The Calculus of Variations

Euler (and later Lagrange) developed the principal ideas for the *Calculus of Variations*. *If* $L(x_1, \ldots, x_n, y_1, \ldots, y_n)$ is a function and we wish to extremise

$$\int_{t_0}^{t_1} L(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t))\, dt$$

along all paths $x(t) = (x_1(t), \ldots, x_n(t))$ that go from $p = x(t_0)$ to $q = x(t_1)$; then this method tells us that we need to solve the equations

$$\frac{\partial L}{\partial x_k}\Big|_{x=x(t), y=\dot{x}(t)} - \sum_{i=1}^{n} \frac{\partial^2 L}{\partial x_i\, \partial y_k}\Big|_{x=x(t), y=\dot{x}(t)} \dot{x}_i(t)$$

$$= \sum_{i=1}^{n} \frac{\partial^2 L}{\partial y_i\, \partial y_k}\Big|_{x=x(t), y=\dot{x}(t)} \ddot{x}_i(t)$$

In our case we want $L = (\sum_{i,j=1}^{n} g_{ij}(x)\, \dot{x}_i \dot{x}_j)^{1/2}$. This is not likely to have a nice form since the same curve may appear in many guises (parametrisations). It turns out that a better way is to extremise the 'energy'

$$E = \int_{t_0}^{t_1} \sum_{i,j=1}^{n} g_{ij}(x(t))\, \dot{x}_i(t)\, \dot{x}_j(t)\, dt$$

for which $L = (\sum_{i,j=1}^{n} g_{ij}(x)\, \dot{x}_i \dot{x}_j)$. The geodesic equations then become

$$\sum_{ij=1}^{n} \left(\frac{\partial g_{ij}}{\partial x_k} - \frac{\partial g_{ik}}{\partial x_j} - \frac{\partial g_{jk}}{\partial x_i} \right) \dot{x}_i\, \dot{x}_j = \sum_{i=1}^{n} g_{ik}\, \ddot{x}_i$$

The paths of shortest length or geodesics are 'straight lines' for our geometry; and one advantage over the method of Gauss is that this definition does not depend on any ambient Euclidean space.

Now Euler had given a way to find the extremal values of such integrals (for the application to Fermat's least time principle; see - box above). Applying his methods one deduces an equation for a path of extremal (minimum or perhaps maximum!) length (see

box on the calculus of variations). Now we make the further 'natural' assumption (which was later overturned in the space-time context by Minkowski and Einstein) that at each point the form $\sum g_{ij} a_i a_j$ is positive for any collection of numbers $(a_1, \ldots, a_n)$. Moreover, the given form can easily be adapted so that $g_{ij} = g_{ji}$; so we assume this. One can then show that the extremal paths are indeed of shortest length (at least locally[5]). Thus the paths of shortest length or geodesics are 'straight lines' for our geometry; and one advantage over the methods of Gauss is that this definition does not depend on any ambient Euclidean space.

[5] See previous footnote 3.

The next question is what quantities need to be determined to determine the geometry uniquely. The functions g_{ij} are clearly sufficient but don't seem to be necessary, since they change depending on the system of coordinates chosen. In a different system of coordinates $(y_1, \ldots, y_n)$, one computes easily (exercise) that the g_{ij} transform into

$$g'_{ij} = \sum g_{kl} \frac{\partial x_k}{\partial y_i} \frac{\partial x_l}{\partial y_j}$$

To take this into account Riemann argues heuristically as follows: The metric is given by the $n(n+1)/2$ functions g_{ij} (not n^2 since the symmetries $g_{ij} = g_{ji}$ have to be taken into account). A change of coordinates is given by n functions (the new coordinate functions). The difference is $n(n-1)/2$ functions which ought to be the number of functions required to determine the geometry. He then proceeds to construct $n(n-1)/2$ functions as candidates. These are the principal curvatures whose definition is outlined below[6].

[6] *A word of warning:* the calculations that follow may seem complicated to the uninitiated reader, even one armed with a paper and pencil. The details can be found in the books listed at the end of the article.

There is a distinguished choice of coordinates around any chosen point o. For each point near 0 let its coordinates be given by the velocity (i.e. direction and speed) with which one must start at o to reach the point in unit time (*Exercise*: What are the coordinates

The Geodesic Normal Coordinates

We write the Taylor expansions of the functions

$$g_{ij}(x) = g^o_{ij} + l_{ij} + q_{ij} + \textbf{higher order terms in the } x_i\textbf{'s}$$

where l_{ij} (respectively q_{ij}) is a linear (respectively quadratic) function of the x_i's. Moreover, by choosing an orthonormal system at the origin o we can assume that g^o_{ij} is 1 if $i = j$ and 0 if $i \neq j$. Now we substitute the given geodesics $(ta_1, \ldots, ta_n)$ into the equations of the geodesic derived in the previous box.

$$\sum_{i,j=1}^{n} \left(\frac{\partial g_{ij}}{\partial x_k} - \frac{\partial g_{ik}}{\partial x_j} - \frac{\partial g_{jk}}{\partial x_i} \right) a_i a_j = 0$$

Equating the coefficients of terms of various orders in t we obtain

$$\sum_{i,j=1}^{n} \left(\frac{\partial l_{ij}}{\partial x_k} - \frac{\partial l_{ik}}{\partial x_j} - \frac{\partial l_{jk}}{\partial x_i} \right) x_i x_j = 0$$

$$\sum_{i,j=1}^{n} \left(\frac{\partial q_{ij}}{\partial x_k} - \frac{\partial q_{ik}}{\partial x_j} - \frac{\partial q_{jk}}{\partial x_i} \right) x_i x_j = 0$$

Now some simple manipulations show that $l_{ij} = 0$. Some more complicated manipulations show that $c_{ij,kl}$ do indeed have the required form of equation (1)

The notion of principal curvature discussed here is *different* from that introduced for surfaces; however the name 'curvature' for these quantities can be justified.

for o?). In other words, the coordinates are so chosen that the paths $(ta_1, \ldots, ta_n)$ are geodesics starting at o for every choice of numbers $(a_1, \ldots, a_n)$ (and for small enough values of t). We substitute this in the equations for the geodesics and perform a simple calculation (see box on geodesic normal coordinates) to obtain the Taylor expansion of the metric in these special coordinates (called *geodesic normal coordinates*),

$$\sum_{i,j=1}^{n} g_{ij}\, dx_i\, dx_j = \sum_{i=1}^{n} dx_i^2 + \sum_{i,j,k,l=1}^{n} c_{ij,kl}\, x_k x_l\, dx_i\, dx_j$$

$$+ \text{ higher order terms in the } x_i\text{'s}$$

We are in a small region of Euclidean space if and only if curvature is zero.

Quadratic Forms

Let $Q(x_1, \ldots, x_n)$ be a quadratic form. We show by induction that there is an orthonormal linear change of coordinates $x_i = \sum_j m_{ij} y_j$ so that

$$Q(x_1, \ldots, x_n) = \sum \lambda_j y_j^2$$

Let S denote the collection of all points $(x_1, \ldots, x_n)$ such that $\sum x_i^2 = 1$. Let $(a_1, \ldots, a_n)$ be a point on S so that $Q(a_1, \ldots, a_n)$ is maximum among all points on S. We can easily show that for any point $(x_1, \ldots, x_n)$ on S such that $\sum x_i a_i = 0$ we have

$$Q(sa_1 + tx_1, \ldots, sa_n + tx_n) = s^2 Q(a_1, \ldots, a_n) + t^2 Q(x_1, \ldots, x_n)$$

Choose an orthonormal change of coordinates so that $(a_1, \ldots, a_n)$ becomes $(1, \ldots, 0)$. The expression we then obtain is

$$Q(x_1, \ldots, x_n) = \lambda_1 y_1^2 + Q'(y_2, \ldots, y_n)$$

where $\lambda_1 = Q(a_1, \ldots, a_n)$ and Q' depends on a smaller number of variables. An induction argument completes the proof.

Moreover, the $c_{ij,kl}$ are numbers such that

$$\sum_{i,j,k,l=1}^{n} c_{ij,kl}\, a_i a_j b_k b_l = \sum_{\substack{1 \le i < j \le n \\ 1 \le k < l \le n}} d_{ij,kl}\, (a_i b_j - a_j b_i)(a_k b_l - a_l b_k) \tag{1}$$

for some numbers $d_{ij,kl}$. The right hand side can be thought of as a quadratic form in the variables $A_{ij} = (a_i b_j - a_j b_i)$ where $1 \le i < j \le n$. This quadratic form can be put into diagonal form by an orthonormal linear substitution $A_{ij} = \sum m_{ij,kl} B_{kl}$ (see box on quadratic forms).

The 'diagonal entries' k_{ij} are then called the *principal curvatures* at o (the minus sign is for historical reasons). This notion of principal curvatures is *different* from that introduced for surfaces in the previous article; however, the name 'curvature' for these quantities is justified as follows. Consider a pair of vectors

$(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$ and the surface described parametrically as the collection of all points $(sa_1 + tb_1, \ldots, sa_n + tb_n)$. The Gaussian curvature of this surface at o can be computed and shown to be

$$\frac{-1}{4} \Big(\sum_{\substack{1 \le i < j \le n \\ 1 \le k < l \le n}} d_{ij,kl} \;\; (a_i b_j - a_j b_i)\,(a_k b_l - a_l b_k) \Big)$$

For manifolds of dimension 2 (i.e. surfaces), this is precisely the Gaussian curvature introduced in the previous article. Now note that there are $n(n-1)/2$ variables A_{ij} and thus there are $n(n-1)/2$ principal curvatures. Thus we have obtained the required $n(n-1)/2$ functions on the manifold; unfortunately the count of Riemann doesn't quite work beyond this point.[7] However, it is true that we are in a small region of Euclidean space if and only if the curvature is zero. It now appears that these depend on a special choice of coordinates, but since this choice was determined by the metric, it turns out that these quantities have an interpretation independent of the choice of coordinates (this was demonstrated by Riemann and later simplified through the work of Christoffel, Cartan and Koszul). Thus we have found a collection of quantities that depends only on the manifold and not on a chosen coordinate system. Moreover, these quantities generalise the notion of curvature and as a by-product we also have a proof of Gauss' *Theorema Egregium*.

Riemann then addressed himself to the question of what this means for space, i.e. the world around us. First of all, this curvature could be computed by making measurements. Though measurements (in his time) had not shown up any curvature, he says this only demonstrates that the curvature is extremely small. Since curvature being exactly zero is precisely the condition for being in Euclidean space, this lays the question of whether space is Euclidean or not, firmly at the door of the experimental sciences.

Since curvature being exactly zero is precisely the condition for being in Euclidean space, this lays the question of whether space is Euclidean or not, firmly at the door of the experimental sciences.

[7] The margin here is too small for a proof!

The remaining problem is that of the global structure of space. Riemann pointed out that the property of unboundedness[8] is the property he called 'extension'. This does not contradict the possibility of the universe being finite. In fact, he points out that if the curvatures of space are everywhere greater than a positive constant, then the universe must close upon itself like a sphere. Thus the global structure of our geometry is directly related to the quantities computed locally, i.e. curvatures.

[8] In the words of an old lady responding to a lecture by Eddington, "if the umbrella has a boundary I could just go to the edge of the universe and poke out my umbrella!"

Summary

Gauss, who was extremely sparing in his praise, came out of Riemann's lecture transfixed with amazement and full of compliments.

It is not at all surprising that Gauss, who was extremely sparing in his praise (he only noted to Bolyai, Lobachevsky, Jacobi, Abel and others that their work was good—since it was in conformity with his own calculations!), came out of Riemann's lecture transfixed with amazement and full of compliments. Most of the details of calculations in this lecture were only available in later papers of Riemann; these appeared late since Riemann was a man with many interests. Other papers by Riemann that were to make a lasting impression were the ones on the theory of functions of a complex variable and the paper on the *Prime Number Theorem* (which led to the famous Riemann hypothesis of number theory). In any case Riemann transformed geometry completely and this explains why the entry of Bernhard Riemann into geometry must always be noted with thunderous applause!

Suggested Reading

J W Milnor. Lectures on Morse Theory, Annals of Math. Studies. No. 55, Princeton University Press, New Jersey, 1963.
Book which deals with many topics covered here rather nicely.

M Spivak. Comprehensive Intoduction to Differential Geometry. vol.II, Publish or Perish, Berkeley, 1970.
The author depended upon the translation of Riemann's lecture given in the above book.

Address for correspondence
Kapil H Paranjape
Indian Statistical Institute,
8th Mile, Mysore Road,
Bangalore 560 059. India

Know Your Chromosomes

3. Hybrid Cells and Human Genetics

Vani Brahmachari

The selective elimination of human chromosomes in mouse-human hybrid cells generates a unique system for cytogenetic analysis. The use of such somatic cell hybrids in chromosome mapping is discussed in this part of the series.

Vani Brahmachari is at the Developmental Biology and Genetics Laboratory at Indian Institute of Science. She is interested in understanding factors other than DNA sequence *per se*, that seem to influence genetic inheritance. She utilizes human genetic disorders and genetically weird insect systems to understand this phenomenon.

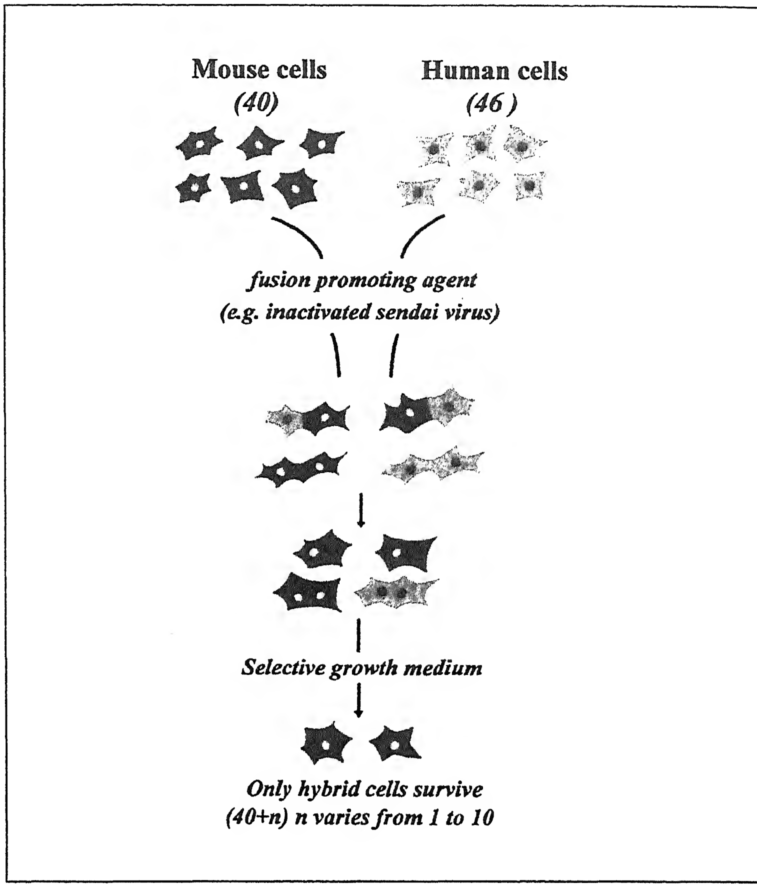
Thanks to our unique abilities we are capable of reigning over almost all living beings: certainly over mice, rats and hamsters. However, the results of artificially fusing a mouse cell with a human cell suggest otherwise. When two cells are fused under suitable conditions, their cytoplasms get mixed first, followed by the fusion of the two nuclei. After nuclear fusion, as the cells continue to divide, some of the human chromosomes are selectively lost while all the mouse chromosomes remain intact. As if compromising on coexistence, a certain number of human chromosomes, varying from one to five pairs, remain stable in these hybrid cells after several divisions. In spite of this chromosome elimination we have exploited hybrid cells to learn about our genetic endowment!!

Exploiting The Hybrids

The basic scheme for generating hybrid cells is shown in *Figure 1*. We have already learnt that human chromosomes can be identified by their staining properties and the specific genes they contain. Therefore it is not difficult to imagine that by careful experimentation one can create a library, not of books, but of hybrid cells. Each clone of cells[1] in the library, would contain only one or a pair of human chromosomes, plus a background of mouse chromosomes. Cell fusions have been carried out not only between human and mouse cells but also between human and

[1] A clone of cells derived from a single parent cell is expected to consist of cells identical in most respects.

Figure 1 The basic scheme for generating hybrid cells.



rat cells. In all these instances, there is selective loss of human chromosomes. The reasons for this are not known. Human cells can be fused among themselves too, but this does not offer any specific advantage in gene mapping. How do hybrid cells help in mapping a gene? Basically one should be able to assess the presence of the human genes either bv their function or by physically identifying them. I will illustrate both of these approaches with a specific example.

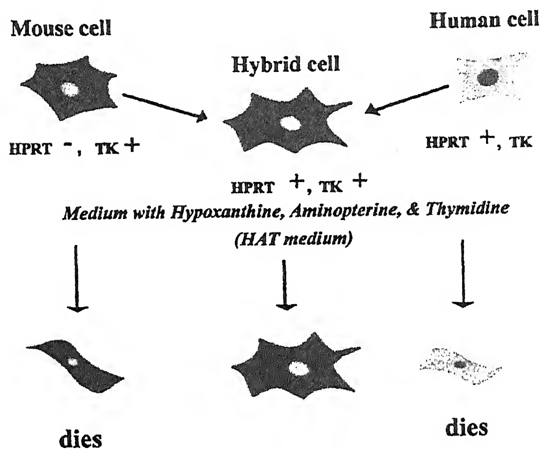
Each clone of cells in the library would contain only one or a pair of human chromosomes, plus a background of mouse chromosomes.

Tracking By Function

Central to this approach is the possibility that the presence of a human chromosome can complement a known defect in the mouse cell. In order to select a hybrid cell of this nature one needs

Selection of Hybrids by Function

Purines like guanine and adenine found in RNA and DNA are synthesized from a combination of simple precursors through several enzymatic reactions. This process is called *de novo* synthesis. Aminopterine (a drug), inhibits *de novo* synthesis. Under such conditions vertebrates utilize a salvage pathway to synthesize nucleotide triphosphates. Two key enzymes of this pathway are hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT) and thymidine kinase (TK).



When aminopterine inhibits *de novo* synthesis, the mouse cells cannot utilize hypoxanthine as they are $HPRT^-$ and similarly human cells cannot utilize thymidine as they are TK^-. But hybrid cells (shown in pink) survive as they have both enzymes; TK from mouse and HPRT from human cells. By karyotyping hybrid cells, it is seen that they have the human X-chromosome. Thus one can conclude that HPRT gene is on the X-chromosome.

to devise a condition in which only the hybrid cells but not the parent cells (namely the mouse and the human cells) survive. This is done by starting with parent cells which are each defective in one of two different enzymes and therefore can survive only in a set of conditions, say growth medium A. But when a hybrid is formed the defects in the two parent cells are compensated or complemented by each other and hence the hybrid can survive in a growth medium B, where the parent cells cannot survive. This is illustrated in the box above.

How do hybrid cells help in mapping a gene? Basically one should be able to assess the presence of the human genes either by their function or by physically identifying them.

Having selected the hybrid cell one can propagate it and analyse its chromosomal profile or karyotype. There is a lucky break here! All mouse chromosomes are *acrocentric*, meaning that the centromere is at one end and they look like the letter 'V' in a

metaphase preparation; this makes them distinct from human chromosomes. We also know (*Resonance*, Vol.1, No.1, January 1996) that one can identify each human chromosome with a specific banding pattern.

Using this approach, one selects for hybrid cells containing the human chromosome bearing the gene that can complement the deficiency in the mouse cell. For instance, mouse cells defective in enzyme E1 and human cells defective in enzyme E2 are chosen as parent cells. Hybrid cells grow in the special growth medium provided they have enzyme E1 coded by the human chromosome along with the complete complement of the mouse genome. Thus one concludes that the human chromosome retained in these hybrid cells has the gene coding for enzyme E1.

With a combination of methods, it is possible to localize a gene not only to a chromosome, but also to a specific band on the chromosomes.

As you can see there are several conditions to be fulfilled before you localize a gene to its chromosome by this approach. The major criteria are (a) availability of parent cells with appropriate deficiencies and (b) a selective medium where only hybrid cells but not the parent cells grow.

Therefore the approach of functional complementation is of limited applicability and has been used to localize enzyme coding genes on chromosomes such as 17, 16, 12 and the X–chromosome. The other approach requires a knowledge of the DNA sequence of at least part of the gene. An analysis based on antibodies specifically directed against the human protein suspected to be expressed by the hybrid cell can also be used for selection.

Mapping by DNA Sequence

Let us assume that we have a DNA fragment of known or unknown function from the human genome. Our aim is to localise this DNA fragment to a specific chromosome. One can have a library of hybrid cells, say from 1 to 23, each retaining one human chromosome. What one does is tag the DNA fragment on

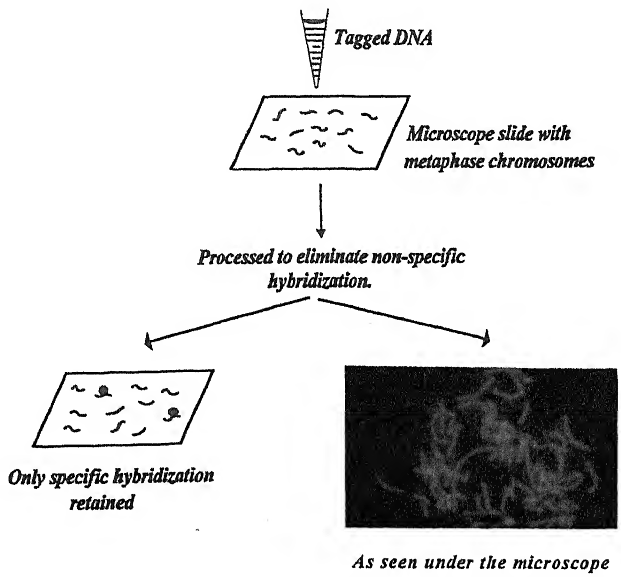


Figure 2 In situ hybridization; the photograph shows the view as seen under the microscope.

hand either with a coloured chemical or a radioactive isotope. This tagged DNA can pair only on the chromosome where an identical DNA stretch is present. This is schematically depicted in *Figure 2*. This process is called *hybridization* and can be carried out either on chromosomes or on DNA derived from the clones. When it is done on a chromosome, it is called *in situ* hybridization. The same procedure can be carried out on chromosomes arrested at the metaphase stage of mitosis, from human cells as well.

Mapping genes by *in situ* hybridization on metaphase chromosomes of hybrid cells has become almost an obligatory step in human genome mapping.

With a combination of methods, it is possible to localize a gene not only to a chromosome, but also to a specific band on the chromosome. For instance, by hybridizing a DNA sequence corresponding to the interleukin (a cell growth factor) coding gene on metaphase chromosomes one can detect hybridization to the long arm of chromosome 4 between band 26 & 27. Thus the map position of the interleukin coding gene is denoted as 4q26-q27. Mapping genes by *in situ* hybridization on metaphase chromosomes of hybrid cells has become almost an obligatory step in human genome mapping.

Table 1: A Listing of Representative Loci Mapped to Chromosomes 2, 3 and 4

Gene/disorder	Chromosomal location	*Mode of inheritance*
1. Apolipoprotein B (APOB)	2p24-p23	*Autosomal dominant*
APOB is the main apolipoprotein of low density lipo proteins (LDL) that occurs in plasma. Deficiency leads to coronary artery disease, gait disturbance, ataxic hand movement.		
2. Colon cancer, familial (FCCI)	2p16	*Autosomal dominant*
The gene in this region is involved in repairing errors in DNA replication. Mutation results in failure of a repair system which leads to DNA instability and colon cancer.		
3. Pulmonary surfactant Apoprotein (PSP-B)	2p12-p11.2	*Autosomal recessive*
The gene codes for a protein associated with lipid rich pulmonary surfactant that prevents lung collapse by lowering surface tension at air-liquid interface. Defect in the gene leads to respiratory failure.		
4. Xeroderma pigmentosum II (XP2)	2q 21	*Autosomal dominant*
The product of this gene is a helicase involved in repairing DNA damage caused by ultraviolet radiation. A defective gene leads to sensitivity to ultraviolet rays and increases the predisposition to skin cancers.		
5. Insulin-dependent diabetes	2q	*Autosomal recessive*
The nature of the gene is not known. Mutations in this region increase the susceptibility to insulin-dependent diabetes melitus.		
6. Brachydactyly Type E (BDE)	2q 37	*Autosomal dominant*
Mutations in this locus lead to short stature, shortening of fingers and reduction in number of digits. The kind and intensity of defects vary between the members of the same affected family. It is a locus which seems to be involved in a complex phenomenon like three dimensional form.		
7. Von Hippel-Lindan syndrome (VHL)	3p26-p25	*Autosomal dominant*
The syndrome is characterised by several carcinomas, renal cysts and hypertension. The nature of the gene(s) is not known.		
8. Hypernephroma (HN)	3p 14.2	*Autosomal dominant*
Mutations at this locus result in hereditary renal cancer, adenocarcinoma of the kidney. The nature of the gene(s) is unknown.		
9. Protein S (PSA)	3p11.1-q11.2	*Autosomal dominant*
It is a vitamin-K dependant plasma protein that prevents blood clotting. The deficiency in protein K results in thrombosis or inappropriate clotting of blood.		

Gene/disorder	Chromosomal location	*Mode of inheritance*
10. Rhodopsin (RHO)	3q21-q24	*Autosomal dominant and recessive forms known*
This is a visual pigment mediating vision in dim light. Defect results in retinitis pigmentosa, defects in retinal pigmentations and night blindness.		
11. Sucrose-isomaltase (SI)	3q25-q26	*Autosomal recessive*
It is an enzyme found in small intestine brush-border membrane, involved in hydrolysing sucrose. Deficiency of the protein results in malabsorption of sucrose from the diet leading to diarrhoea, disaccharide intolerance, and kidney stones.		
12. Huntington Chorea (HD)	4p 16.3	*Autosomal dominant*
The disease gives rise to progressive, selective neural cell death associated with choreic movements and dementia. It is associated with CAG triplet repeat expansion in a gene called huntingtin.		
13. Cyclic nucleotide gated channel Photoreceptor (CNCG)	4p14	*Autosomal dominant and recessive forms known*
Involved in the function of rods and cones in the eyes. Defective protein leads to retinitis pigmentosa.		
14. Dysalbuminemia (DALB)	4q11-q13	*Autosomal dominant*
Gene codes for albumin which is one of the most abundant proteins of blood serum. It acts as a carrier for steroids, fatty acids and thyroid hormones. Mutations in this gene result in disorders of connective tissue like cartilage, tendon and ligament.		
15. Mucolipidosis II (ML2)	4q 21-q23	*Autosomal recessive*
The disorder is characterised by congenital dislocation of the hip, thoracic deformities, hernia, slower psychomotor development and restricted joint movements. It is suspected that there is leakage of enzymes from lysosomes, the suicide bags of the cell.		
16. Interleukin-2 (IL2)	4q 26-q27	*Autosomal recessive*
It is a cell-growth factor required for the proliferation of lymphocytes. Defect in the gene results in severe combined immunodeficiency.		
17. Muscular dystrophy Facio scapula humeral (FSHD)	4q 35-qter	*Autosomal dominant*
The disorder leads to muscle weakness; symptoms appear early in infancy first in the face, upper arms and shoulder muscle. Gene responsible not identified.		

Genes on Chromosomes 2, 3 and 4

A representation of chromosomes 2, 3 and 4 is shown *in Figure 3*. As of now the total number of genes localized to each of these is: 199 (chromosome 2), 191 (chromosome 3) and 150 (chromosome 4). These genes include those responsible for encoding enzymes, growth factors and proteins involved in neural pathways. *Table 1* lists some genes from chromosomes 2, 3 and 4. As one may notice, there is no relationship between the chromosomal location of genes and their function. In most cases, there is no clustering of genes just because they are part of the same metabolic pathway.

Huntington's disease is an example in which the severity of the disease depends on the sex of the parent from whom the defective gene is inherited.

One of the genes mapped to chromosome 4 is the gene for Huntington's disease or Huntington's chorea. Named after George Huntington, a physician who described the disorder in 1872, it is a dominant autosomal disorder that leads to nerve cell death, progresses with age and is associated with rigidity, loss of memory and personality changes. Typically, the patients die 10-15 years after the onset of the disease. This disorder is representative of a class of disorders which may not be seen at birth, but occurs at different ages in different patients. The age of onset of the disease can be from 10 to 70 years. This is also an example in which the

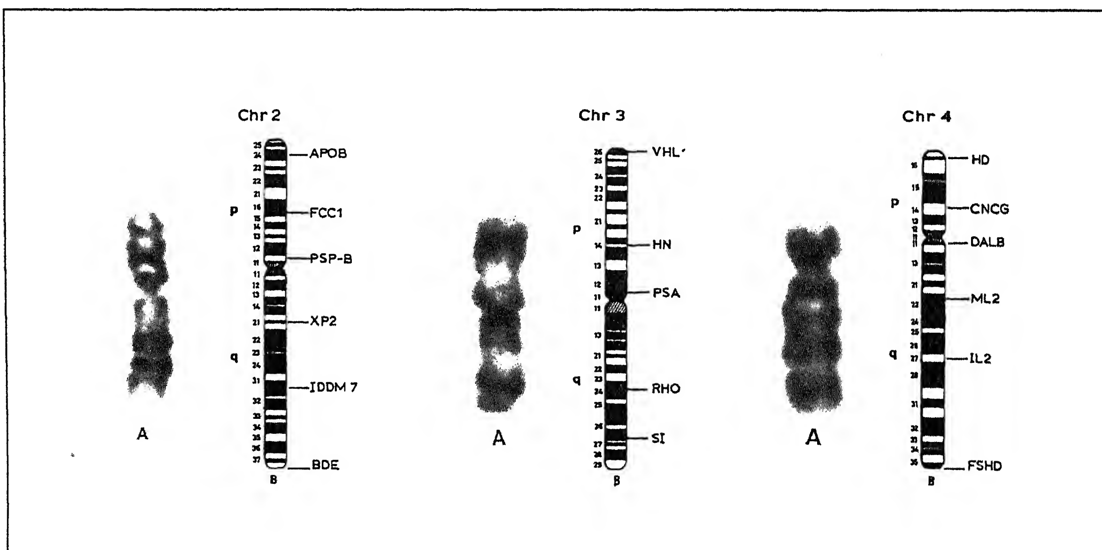


***Figure 3** A representation of chromosomes 2, 3 and 4.*

severity of the disease depends on the sex of the parent from whom the defective gene is inherited. If the child inherits the defective gene from the father, it is likely to have a more severe disorder than the father and at an age earlier than him. When inherited from the mother, both severity and age of onset are likely to be simillar between the child and the mother.

Address for correspondence
Vani Brahmachari,
Developmental Biology and Genetics Laboratory,
Indian Institute of Science,
Bangalore 560 012, India.

The identification of the gene responsible for Huntington's disease was announced in 1993 in a research paper authored by 58 scientists belonging to six different groups! The protein coded by the gene is named 'huntingtin' and is believed to exert its effects by interacting with other proteins. The nature of the mutation that leads to the disorder has helped us understand at least partially, the basis of differences in severity and age of onset from one generation to the other. But how the human system tolerates the absence of a functional gene product in early life but not later is far from clear. Perhaps there is functional redundancy suggesting that nature, the excellent designer, has provided sufficient backup to avoid a system breakdown.

Suggested Reading

Nils R Rigertz and Robert E Savage. Cell Hybrids. Academic Press. New York, San Francisco, London. 1976.

Daniel Hartl. Human Genetics. Harper and Row Publishers. New York, Cambridge, London. 1983.

Friedrich Vogel and Arno G Motulsky. Human Genetics: Problems and Approaches. Springer - Verlag. Berlin, Heidelberg, New York, Tokyo. 1986.

Victor A McKusick. Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive and X linked Phenotypes. Vol I and II Tenth Edition. The Johns Hopkins University Press. Baltimore and London. 1992.

Anonymous poetic supplication ...

Grant, Oh God, thy benedictions
On my theory's predictions,
Lest the facts when verified,
Show Thy servant to have lied.

What Happened to Hyakutake?

On the Trail of a Comet

B S Shylaja

B S Shylaja is with the Bangalore Association for Science Education, which administers the Jawaharlal Nehru Planetarium as well as a Science Centre. After obtaining her Ph.D. on Wolf-Rayet stars from the Indian Institute of Astrophysics, she continued her research on novae, peculiar stars and comets. She is also actively engaged in teaching and writing essays on scientific topics to reach students.

The comet Hyakutake was discovered on January 30th of this year. In this article the author describes many special features associated with it.

The beauty of a particular celestial object attracted a lot of attention in March 96. The hero was not the brilliant (eclipsed!) sun, but a tiny object called comet *Hyakutake*.

This comet, discovered on the 30th of January by the Japanese astronomer Yuji Hyakutake, has many special features. When it was first spotted with binoculars, it was a fuzzy blub of magnitude [1] 11. However a quick follow up on the position details revealed information on the orbit and an extraordinary situation arising as a consequence.

Generally, all the comets are discovered, i.e. detected for the first time in their orbit, at distances of about 2-3 AU (one astronomical unit is the mean distance from the earth to the sun, about 150,000,000 km.) as very faint hazy blobs. Most of them go unheard of (by the layperson) since they do not reach the visibility limit of even small telescopes. They become brightest when they are very close to the sun, that is at the *perihelion*. This obviously indicates that they are available above the horizon in the absence of sunlight for a very limited period. We did experience this 10 years ago, when comet *Halley* was back to salute the sun. Twilight and the larger air mass through which the object had to be viewed, diminished observations for any scientific purpose. The same is true with any other comet, be it *Giocobinni-Zinner* or *Austin* or *Levy*. Therefore for astronomers to be able to see a comet at a very comfortable elevation (i.e. high above the

[1] Magnitude six is about the faintest that the naked eye can see. Each additional magnitude makes the object about a factor of 2.5 dimmer, so magnitude 11 is a hundred times too dim to be seen unaided!

horizon) is a luxury and *Hyakutake* offered this very generously. How did it do that?

It wandered slowly in the constellation of Libra for almost the whole of February and most of March. Rising just about midnight, it crossed the meridian by the early hours of the day making life very easy for telescope users. This guaranteed 3-4 hours of astronomical observation in the clear winter months. By the second week of February, the spectroscopic details started pouring in, in the IAU circulars. (These are quick messages that are despatched for the benefit of the astronomical community.)

It is a luxury for astronomers to be able to see a comet at a very comfortable elevation (i.e. high above the horizon) and *Hyakutake* offered this luxury very generously.

The comet then headed towards the sun in its usual course. Surprisingly one the way to its perihelion point, it crossed us via the north pole and as a consequence was available above the horizon all through the night for about 4 days in the last week of March. The distance was a mere 0.1 AU on March 25th and it showed up quite brightly letting millions enjoy this rare sight of a comet at a pole. (In 1983, the comet *IRAS-Araki-Alcock* provided a similar opportunity, but it enacted this at very short notice depriving many of the rare sight.) The tail, which was faintly recognizable till March 20th, grew and then changed its orientation by the end of the month.

Astronomers noticed one more peculiarity of the comet on March 26th just after its closest approach to the earth. The nucleus of the comet which appeared very sharp to the naked eye, had split into two fragments. The two drifted slowly apart and their velocities were also measured. We are now immediately reminded of the "string of pearls" which created exciting news a couple of years ago. The comet *Shoemaker-Levy 9* which plunged into Jupiter had broken into 21 pieces. Why do comets break into pieces? Before answering this question we should know what comets are.

What are Comets?

Comets are very tiny (average size 4 km) fragile, irregularly

The Rocket Effect

Imagine that there is a snowman on the surface of a rotating comet. The snowman experiences a sunrise, sun at zenith and a sunset, akin to his counterpart on earth. Therefore the rate of melting of the snow increases gradually from morning, reaches a maximum by about mid day and reduces by evening. The chillness in the night however is sufficient to freeze him back again. After several days we would notice the evaporated material accumulate over the head of the snowman, because of the maximum rate of melting at midday. Some of this material will escape from the comet, because of the low escape velocity and this will generate a recoil effect. The direction of the recoil that is whether it is clockwise or anticlockwise will be decided by the spin of the comet. This effectively reduces or increases the orbital velocity and is called the *rocket effect*. This rocket effect was suggested as early as 1835, when Bessel observed the apparition of *Halley* that year. Comet *Halley* was spinning the 'right' way and got delayed in reaching the perihelion point at successive apparitions. Comet *Encke* was spinning the 'wrong' way and reached perihelion ahead of the scheduled time. The rocket effect successfully explained these anomalies.

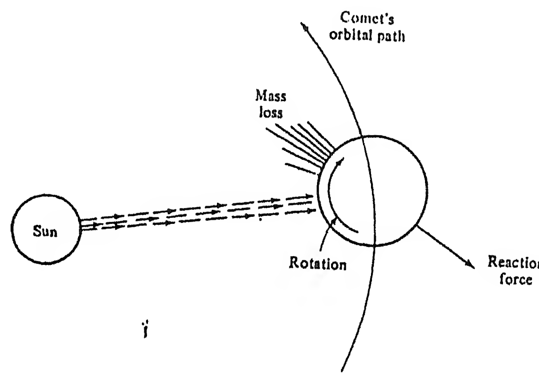


Figure 1 As a comet approaches the sun, the gas molecules evaporate from the surface of the nucleus. The ejected gases have a preferential direction with reference to the sun-comet line. The recoil of the nucleus due to this gets the name 'rocket effect'.

shaped bodies generally referred to as dirty 'snowballs', by virtue of their various ingredients. Most of them are visible because the volatile gases vaporise due to the sun's heat energy. This increases the density of the surrounding gaseous material as they move closer. The gases envelope the central body called the nucleus and at the perihelion point they block the view of the nucleus.

The eccentricity, e, is a measure of the departure of an ellipse from a circle. A circle has eccentricity zero, while the parabola has 1.

Comets obey the same universal laws of motion as the planets, but unlike planets they have elongated orbits. Mathematically, we can say their eccentricities are close to unity. The eccentricity, e, is a measure of the departure of an ellipse from a circle. A circle has eccentricity zero, while the parabola has 1. All the planets

have eccentricity e close to 0, with only Mercury and Pluto deviating to about 0.1. That explains why the orbits of periodic comets resemble a parabola, even though they are really ellipses.

Many comets get perturbed very easily by a planet on their way to the perihelion. The comet *Halley*, known for its periodic visits, thus has a period ranging from 74.4 years to 79.2 years. In the original paper in 1709 by Edmund Halley, which predicted the return of the comet, he mentions that Jupiter undoubtedly had serious effects on the comet's motion.

It must be remembered that comets are fragile bodies. Their tensile strength is very low — of the order of 1000 dynes/cm^2 (Compare this with the strength of, say aluminium, which is of the order of 10^9 dynes/cm^2.) That is similar to a snowball which can be pulled apart with bare hands. The gravitational pull, by virtue of a comet's low mass, is also very small. Remember that it requires a velocity of 11 kms/sec for an object to leave the earth, whereas a similar escape velocity for a typical comet is about 1metre/sec. What is the consequence of this low surface gravity?

Comets obey the same universal laws of motion as planets, but have elongated orbits. Mathematically, we can say their eccentricities are close to unity.

We learned that the material which was frozen in the nucleus, evaporates during the period of perihelion approach. The evaporated dust and molecules are dragged along for a while and ultimately become independent of the nucleus. Moreover, they are pushed into the shadow region of the nucleus which is away from the sun. The solar wind pushes and aligns them neatly as the tail of the comet.

As the comet approaches the perihelion point, the molecules are exposed to the intense ultraviolet radiation, which breaks them apart and they become identifiable by their spectra. Very close to the sun, they are deprived of their electrons as well and the ions thus formed are electrically charged and are exposed to the magnetic lines of force of the solar wind. They align themselves into a different tail identified as the ion tail of the comet.

Models in Cometary Astronomy

1950 was the most important year in cometary astronomy. Oort proposed the 'cloud' or a store house of comets in the remote reaches of the solar system, a hundred times further than Pluto. Kuiper suggested a 'belt' beyond the orbit of Neptune for short period comets; Whipple offered the 'dirty snowball' model for the cometary nucleus.

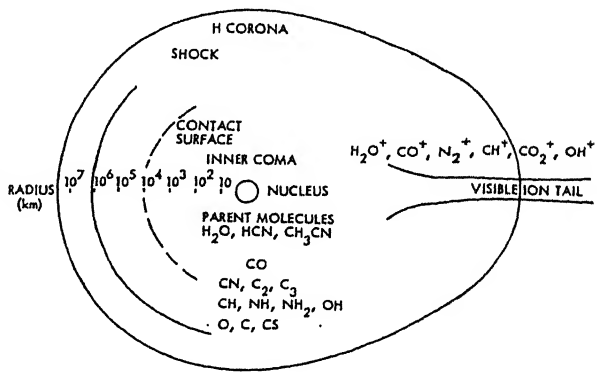


***Figure 2** The general distribution of the ions and molecules in a cometary coma shows the parent molecules like H_2O, HCN and CH_3CN in the inner coma. As the comet nears the sun, other molecules and ions are created.*

The model predicts the rate of production of molecules at different heliocentric distances. The gaseous activity becomes recognizable at about 3 AU. Beyond this distance the spectrum shows only continuum. At 3 AU the CN emissions appear, at about 2 AU C_3 and SH_2 appear. As the comet proceeds towards the sun, at about 1.5 AU the spectral lines of C_2, CH, OH and NH appear. By now the continuum is very weak or in other words, the emissions dominate the spectrum. These strong colourful lines look beautiful through a spectrograph. At still closer distances, ionised species like CO^+, OH, N^+_2 and CH^+ appear. The study of the spectrum and its variation through the perihelion passage is important, because differences in the spectra reveal the nature of the nucleus itself.

Comets are endowed with the special privilege of developing tails millions of kilometres long, that can be shaped, or stretched to 25 to 30 degrees of the sky.

Thus comets are endowed with the special privilege of developing tails millions of kilometres long, that can be shaped, or stretched to 25 to 30 degrees of the sky, or can be colourful, with sodium providing a yellow tint. All this beauty is achieved at the cost of particles in the tail that are lost forever!

The Breakup of the Nucleus

Although the fragmentation of the nucleus was best publicized in the *Shoemaker-Levy* episode, more than 25 such cases are known. One recent event was *McHoltz* which split in 1995. The splitting

of *West* in 1977 was also well studied. Sometimes two or more comets share the same orbit indicating that they were once a single piece. Calculations point to Jupiter as responsible for this type of splitting in many cases.

When comets split, the pieces do not fly apart.

There is a family of about 16 comets, generally referred to as the *Kreutz* family (after Heinrich Kreutz, who investigated the comets and identified this group), which passes very close to the sun—within 3 million kms or less. Marsden of the International Bureau of Telegrams suggests, based on orbit computations, that all these may have resulted from the fragmentation of a comet which appeared in AD 1100, which in turn may have been part of a great comet of BC 372. Half of these could not survive the perihelion passage and splashed onto the sun. The other members of the family survived, but broke into pieces.

When comets split, the pieces do not fly apart. Calculations assuming the mean density to be about 0.5 gm/cc, (half of that of water), and a rotation period of 12 hours for a comet of size 12 km show that a particle at its equator (if it has any defined), would depart with only a velocity of 0.72 meters/sec relative to the centre of the comet. Many comets rotate much more slowly. This means that the drifting of pieces is not observable. Even when a piece is freed of the gravitation of the nucleus, it drifts slowly with continued weak interactions, which can cause further splitting of the fragments.

The pieces of *Hyakutake* which now have a relative velocity of about 0.8 km/s are being monitored closely. The two fragments have separated. This could have given us the spectacle of two independent tails, as developed by comet *Biela* 150 years ago. It appears, however, that this has not happened.

Address for correspondence
B S Shylaja
Jawaharlal Nehru
Planetarium
High Grounds
T Chowdaiah Road
Bangalore 560 001, India.

Suggested Reading

R A Lyttleton. The Comets and their Origin. Cambridge University Press. 1953.
Nigel Calder. The Comet is Coming. BBC. 1980.
J C Brandt and R D Chapman. Introduction to Comets. Cambridge University Press. 1981.
Various circulars from the IAU.

The Changing Forms of Comet Hyakutake (1996-B2)

Pictures of comet given below courtesy T Chandrasekhar, N M Ashok and J N Desai (PRL)

27 March 1996
3.15 a.m.

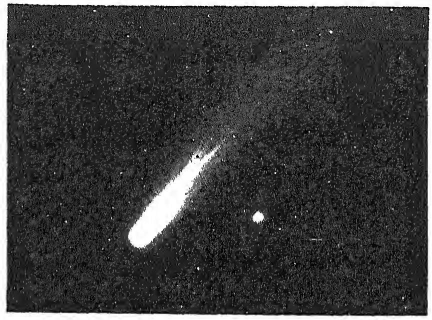
A spectacular photograph of *Comet Hyakutake* taken during its close approach to the earth of approximately 15 million km. A telephoto (focal length = 105mm, f/2.8) camera mounted on the drive of the large 1.2m telescope was used. The exposure time was 30 minutes. During the exposure the cometary motion was frozen by tracking it using a guiding telescope. The motion of stars in the field relative to the comet show up as star trails (smaller linear structures). The exposure was made when the comet was at a distance of 0.10 AU from the earth and 1.03 AU from the sun (1 AU = 150 million km). The photograph covers a field of view 6 degrees x 4 degrees. The tail is seen to extend at least 5 degrees.

5 April 1996
8.15 p.m.

Schmidt photograph with an exposure time of 14 minutes, 30 seconds. A central ray extends to the edge of the field and beyond (>3 degrees). The star seen near the cometary head is *K Persei* (visual magnitude ~ 3.8). The comet was at a distance of 0.37 AU from the earth and 0.81 AU from the sun. The field of view of the photograph is 2.8 degrees x 2.4 degrees.

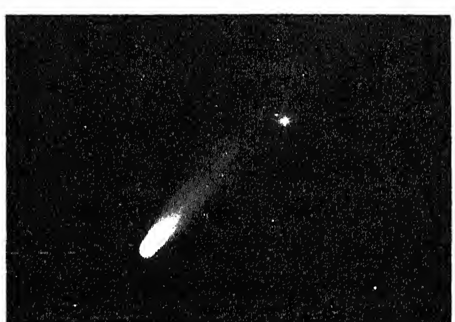
l photographs have been taken from Gurushikhar Observatory, Mt.Abu which is a field station of Physical esearch Laboratory (PRL), Ahmedabad. The observatory is at an altitude of 1680m above sea level.

9 April 1996
8.30 p.m.



Another spectacular view of Comet *Hyakutake* showing the intricate structures in the tail of the comet. The exposure time was 10 minutes. The bright star to the right is the famous eclipsing binary star *Algol* (visual magnitude ~ 2.1). The angular separation between the cometary nucleus and the star is about a degree. The prominent ray structure and kinks along the tail are clearly seen. The tail extends at least 2.5 degrees. At the time of this exposure, the comet was at a distance of 0.49 AU from the earth and 0.71 AU from the sun. The field of view of the photograph is 3.4 degrees x 2.4 degrees.

10 April 1996
8.30 p.m.



Schmidt photograph with an exposure time of 10 minutes. The bright star seen in the tail region is the famous eclipsing binary star *Algol* (visual magnitude ~ 3). The angular separation between *Algol* and the cometary nucleus is 1.43 degrees. The fainter portions of the tail extend beyond the scale of the photograph (>2.5 degrees). At the time of this exposure, the comet was at a distance of 0.53 AU from the earth and 0.69 AU from the sun. The field of view of the photograph is 3.4 degrees x 2.4 degrees.

Colin S Pittendrigh: An Appreciation

The Life of a Darwinian Clock-Watcher

L Geetha

L Geetha studies circadian rhythms in honeybees, mice and humans.

A tribute to Colin S Pittendrigh, one of the three musketeers of chronobiology, who passed away on 20 March 1996.

It was fairly recently that I discovered CBT on the internet. CBT (Centre for Biological Timing), is an organisation that provides, among other things, information about research in the field of chronobiology. And so it was with great expectation of various kinds of interesting details about biological rhythms that I subscribed to it. It is my misfortune that the first piece of information that I received through CBT was that of the demise of one of the doyens of chronobiology, Colin S Pittendrigh. Nature could have been a little kinder to me. While I passed this sad news on to others in the field, it occurred to me that perhaps I should write a few words about him.

It required a lot of courage to speak up and prove to the world that clocks in fact exist and organisms exhibit rhythmic phenomena owing to the presence of internal biological clocks.

An introduction to the field of chronobiology to any student entails an introduction to the three musketeers, E Buenning, C Pittendrigh and J Aschoff. The first seminar that many students give is on one of the fascinating papers of Pittendrigh. It was thus, that I was introduced to him and the multitude of his research papers. Most of the intelligent experiments that have been conceived in this field were his brain work. At this moment, all of us are aware of the existence of rhythmic phenomena and we accept the presence of internal clocks. But this was not the case a few decades ago. Indeed, the attribution of the closing and opening of leaves in plants to the presence of an internal clock which measured external time, was often ridiculed. It required a lot of courage to speak up and prove to the world that clocks in fact exist and organisms exhibit rhythmic phenomena owing to the presence of internal biological clocks. Well, there were a handful

of them who had this courage and their undaunting efforts to express themselves strongly has brought forth a slow transition from non-belief to belief !

Colin Stephenson Pittendrigh (1918 - 1996)

Pittendrigh's early research papers provided lucid explanations of the fundamental concepts of chronobiology such as entrainment and free-run. He, along with Aschoff introduced the 'oscillator language' to explain the mechanisms of entrainment and made innumerable predictions of how these overt behaviours by organisms could be the end product of the functioning of the internal oscillator. Pittendrigh also predicted that the clock should be sensitive to light and that this sensitivity should be different at different times of the day. To prove this, he constructed what is called the *phase response curve (PRC)* which describes the sensitivity of animals to light pulses at various phases. Pittendrigh describes the PRC poetically as the "footprint, as it were, of the pacemaker's (or oscillator or clock) time course"! He also demonstrated by an elegant set of experiments that the shifts in the phases caused by light pulses should be instantaneous. These experiments resulted in a series of four landmark papers along with S Daan in the *Journal of Comparative Physiology* (1976). The seeds of most modern experimentation are contained in these papers.

Pittendrigh had an open mind for new facts and was willing to change his views given sufficient experimental evidence. Thus, in his early experiments he talked about "the oscillator" inside biological systems and how "it" brings forth rhythmic expressions. Later, his own experiments on squirrels proved that there may be more than one oscillator which controls the activity rhythm of these animals. Thus he attempted to explain the occurrence of transients in the *Drosophila* system by a master oscillator and a slave oscillator. Now it is believed that multicellular organisms possess multiple oscillators, each controlling particular activities and still managing to remain in synchrony with each other. And according to Pittendrigh "it is a plausible and adequate

Pittendrigh had an open mind for new facts and was willing to change his views given sufficient experimental evidence

The greatness of a scientist lies not only in performing good experiments but also in reporting them to the world, in a convincing manner. Pittendrigh did this most effectively — each of his papers is a classic in itself.

interpretation for multicellular organisms".

The fruit fly *Drosophila* has now turned out to be a model system for most of the genetic and molecular work on clocks. Pittendrigh elevated *Drosophila* to a similar status in chronobiology. He demonstrated for the first time that biological clocks are temperature compensated i.e. clocks operate normally irrespective of the external temperature. Later, he proved such temperature compensation also in the unicellular *Euglena*. He extended it to *Neurospora*, and introduced it to the world "as a system in which the genetics of the clock could be pursued". Today, both *Neurospora* and *Drosophila* have indeed become powerful tools in the study of genetics of clocks. At this point of time, it has been established that the genes *per* in *Drosophila* and *frq* in *Neurospora* are responsible for the clock activities! They have been cloned, sequenced and exceptionally good progress has been made in deciphering the nature of their operation.

Address for correspondence
L Geetha
Centre for Ecological Sciences
Indian Institute of Science
Bangalore 560 012, India.

The greatness of a scientist lies not only in performing good experiments but also in reporting them to the world, in a convincing manner. Pittendrigh did this most effectively- each of his papers is a classic in itself. His research papers are written in such a simple, lucid and logical manner that even a beginner can feel proud that he understands them. I consider his "Reflections of a Darwinian Clock-Watcher" as his *magnum opus*, a brilliant, anecdotal article, written at the age of seventy five! It is a tragedy that he is no longer among us physically and it is indeed a great loss for the field of chronobiology! But the spirit of Colin Pittendrigh will live for ever and be rekindled by the writings of many generations of new students the world over.

Suggested Reading

Hickory Dickory Dock. T R Raghunand. *Resonance* 4:76. April 1996.

Stephen J Gould's view ... Contingency is rich and fascinating; it embodies an exquisite tension between the power of individuals to modify history and the intelligible limits set by laws of nature. The details of individual and species's lives are not mere frills, without power to shape the large-scale course of events, but particulars that can alter entire futures, profoundly and forever. (from *Eight Little Piggies*)

Energy Storage and Retrieval

The Secondary Battery Route

A K Shukla and P Vishnu Kamath

Harnessing sunlight for the production of electrical energy is an engrossing prospect. The crucial concept underlying the success of solar power stations is energy storage and its retrieval on demand which can be most effectively achieved with storage batteries. This article highlights the chemistry of existing and emerging battery technologies.

A K Shukla is at the Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore. His research interests are in the area of materials electrochemistry with special emphasis on batteries, fuel cells and solid ionics.

P Vishnu Kamath is at the Department of Chemistry, Central College, Bangalore University, Bangalore. His research interests are in the area of materials chemistry with special emphasis on electrode materials. He is also actively involved with various environmental groups in Karnataka.

Technological development in this century has been characterized not only by the increasing consumption of energy but also by the emergence of hydrocarbons as the primary source. The process of development is threatened by the limited reserves of coal and oil. In addition, the deleterious effects of excessive consumption of hydrocarbons on the economy and ecology of a large part of the world is too well known to be recounted here. These have brought into sharp focus the need for developing new environmentally benign *non-conventional* or what we would prefer to call *alternative energy sources*.

Wind, solar and tidal energies are available in almost all parts of the globe and the efficacious harvesting of these energy sources will also alleviate the problems associated with energy transmission and distribution. However these energy sources are intermittent and exhibit annual, seasonal as well as diurnal variations. They are available at certain times of the day or year and not available at other times. The key to the successful utilisation of these energy sources lies in the development of suitable energy storage devices.

The key to the successful utilisation of these energy sources lies in the development of suitable energy storage devices.

Among the many energy storage devices presently in vogue, batteries are the most common. Batteries are electrochemical devices which convert chemical energy into electrical energy.

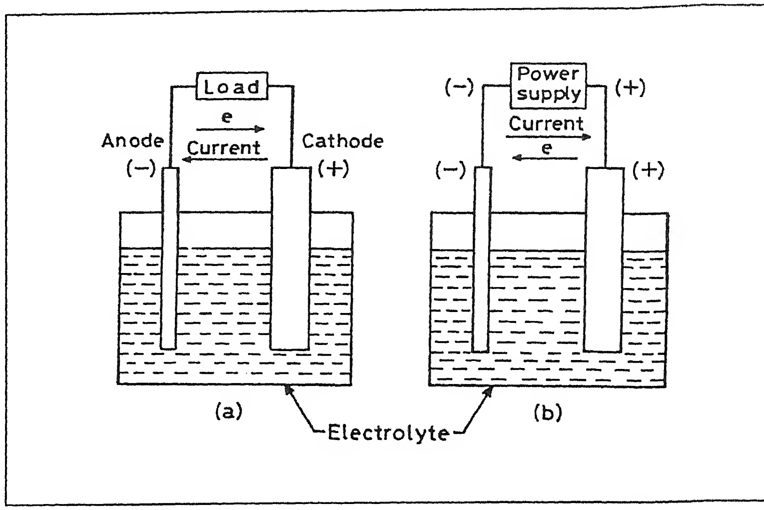


Figure 1 The discharge (a) and charge (b) processes in a battery.

That is, they use chemical reactions to produce electricity. Such electrochemical devices are called *galvanic cells* in contrast to *electrolytic cells* which consume electrical energy to bring about a chemical reaction.

A battery consists of two electrodes, an anode and a cathode, and an electrolyte through which electrically charged particles can move (*Figure 1*). Two chemical reactions take place at the same time. The reaction taking place at the anode is an oxidation reaction of the type $R \rightarrow O + ne$, where R is the reduced species and O, the oxidised species. Such a reaction leads to an excess of electrons, ne, at the anode. It is also called the negative (–ve) electrode. The chemical reaction taking place at the cathode is a reduction reaction of the type $O' + ne \rightarrow R'$, which results in a depletion of electrons. Therefore the cathode is also called the positive (+ve) electrode. For example in a nickel-cadmium battery, Cd is oxidised to Cd^{2+} at the anode and Ni^{3+} is reduced to Ni^{2+} at the cathode and the reactions can be represented as

$$Cd \rightarrow Cd^{2+} + 2e \quad \text{at anode (–ve electrode),}$$
$$Ni^{3+} + e \rightarrow Ni^{2+} \quad \text{at cathode (+ve electrode).}$$

When the battery is connected to an external circuit (load), the excess electrons from the anode flow through the circuit and back

Table 1. Desirable features in a battery

Battery performance parameter	Definition	Desired target
Energy density	Energy (stored) per kilogram of battery weight (Wh kg^{-1}) or per litre of battery volume (Wh l^{-1})	HIGH
Power density	Ratio of power available from a battery to its weight (W kg^{-1}) or volume (W l^{-1})	HIGH
Self discharge	Loss of charge due to parasitic reactions between periods of use	LOW
Internal resistance	Sum of electrical (ionic and electronic) resistances of the battery components	LOW
Cycle life	Number of charge-discharge cycles over which the energy density can be sustained under specific conditions (applicable to a secondary battery)	HIGH
Efficiency	Ratio of the charge output to the charge input during a charge-discharge cycle for a secondary battery	HIGH

to the cathode. As the electrons move through the circuit they lose energy. This energy may be used to create heat or light as in an electrical heater or light bulb, or to do work as in a motor. The flow of electrons results in a *current* and by convention the direction of flow of current is taken as opposite to the direction of flow of electrons. The energy released per unit charge while the current flows through the circuit is called *voltage*. The product of the current and the voltage is the *power* delivered to the circuit. When a battery delivers electric current to an external load, certain active materials in the battery are chemically converted into other materials at lower energy states and the battery is eventually fully discharged.

The battery most commonly known to us is the dry cell, which we use in our transistor sets or torches. These are purchased in their charged state and discharged through use and then discarded.

Table 2. Target characteristics for a high performance battery

Performance parameter	Long life type	High energy density type
Weight energy density (Wh kg^{-1})	120	180
Wolume energy density (Wh l^{-1})	240	360
Cycle life	3500	500
Efficiency (%)	>90	>85

Others: Environmental stability, safety, easy maintenance, high range of operational temperature (–20 to 50^0 C), compactness, ruggedness and low cost.

Primary and Secondary Batteries

The battery most commonly known to us is the dry cell, which we use in our transistor sets or torches. These are purchased in their charged state and discharged through use and then discarded. Such cells are known as *primary cells*. They are of limited use as they deliver much less energy than what is required to construct them. Besides, the growing need for recycling resources requires that the discharged battery should be reusable a large number of times. In other words, we look for a *secondary* (rechargeable) or storage battery with a long *cycle life*. A secondary cell after discharge can be recharged by passing electric current through it in the reverse direction (*Figure 1*). During recharge it behaves like an electrolytic cell and converts electrical energy into chemical energy and the active electrode material is retrieved.

A high performance battery should have a maximum energy density at an optimum power density (rate of discharge), minimum internal resistance, maximum charge retention, mechanical strength and a long cycle life(see *Table 1*). The target values for all these parameters which would help define a high performance battery are given in *Table 2*.

A high performance battery should have a maximum energy density at an optimum power density (rate of discharge), minimum internal resistance, maximum charge retention, mechanical strength and a long cycle life.

How Do We Construct a Battery?

A battery basically consists of two electrodes, at each of which a chemical reaction takes place. Every chemical reaction is associated with a certain free energy change, ΔG^0 (under standard conditions), which can be characterised by a potential E^0, such that $\Delta G^0 = -nFE^0$ (where n is the number of electrons involved in the reaction and F, the Faraday constant). This potential is called the single electrode potential. In a battery, the single electrode potentials of the two electrodes differ from one another so that electrons are released at the anode at a high energy. The electrons pass through a load and lose energy and are consumed at the cathode at a lower energy.

From the 100 odd elements known to us in the periodic table, nearly 5000 (i.e., 100´99/2. The number is larger if variable oxidation states of some elements are taken into account.) pairwise combinations of single electrode reactions involving stable reactants and products can be theoretically envisaged. This leads to a similar number of different possible electrochemical energy storage systems. However in practice, more than a century of effort in the development of batteries has resulted in only a few systems of practical importance.

The small number of successful systems compared to the large number possible in theory suggests that a workable electrochemical energy storage system is critically dependent on several factors. Two obvious factors are cost and availability of required materials. In addition, a few factors related to the chemistry of the electrode materials play an important role.

The small number of successful systems compared to the large number possible in theory suggests that a workable electrochemical energy storage system is critically dependent on several factors. Two obvious factors are cost and availability of required materials.

Chemistry of Reversible Electrodes

The electrode materials in a rechargeable or secondary battery must undergo a reversible chemical reaction. A typical electrode reaction can be schematically written as

$$M^{z+} \rightleftharpoons M^{(z+\Delta)+} + \Delta e \qquad \text{(I)}$$

Table 3. Electrode reactions in different batteries. E^0 values have been given where they are known. Arrows pointing to the right (→) correspond to the discharge reactions and arrows pointing to the left (←) correspond to the charging reactions. Advantages and disadvantages of the batteries are also given.

We must develop new batteries with enhanced performance characteristics for communication, space, automotive, and traction purposes.

Table 3. Electrode reactions in different batteries.

- Lead acid battery

$PbO_2+4H^++SO_4^{2-}+2e \rightleftharpoons PbSO_4+2H_2O$ $E^0=1.69$ V (cathode)

$Pb+SO_4^{2-} \rightleftharpoons PbSO_4+2e$ $E^0=-0.36$ V (anode)

$Pb+PbO_2+2H_2SO_4 \rightleftharpoons 2PbSO_4+2H_2O$ $E_{cell}=2.05$ V (net)

Low energy density, high Pb toxicity, corrosive but rugged

- Nickel-iron battery

$2NiOOH+2H_2O+2e \rightleftharpoons 2Ni(OH)_2+2OH^-$ $E^0=0.49$ V (cathode)

$Fe+2OH^- \rightleftharpoons 2Fe(OH)_2+2e$ $E^0=-0.88$ V (anode)

$2NiOOH+Fe+2H_2O \rightleftharpoons 2Ni(OH)_2+Fe(OH)_2$ $E_{cell}=1.37$ V (net)

Poor performance of Fe electrode, not maintenance-free

- Nickel-cadmium battery

$2NiOOH+2H_2O+2e \rightleftharpoons 2Ni(OH)_2+2OH^-$ $E^0=0.49$ V (cathode)

$Cd+2OH^- \rightleftharpoons Cd(OH)_2+2e$ $E^0=-0.76$ V (anode)

$2NiOOH+Cd+2H_2O \rightleftharpoons 2Ni(OH)_2+Cd(OH)_2$ $E_{cell}=1.25$ V (net)

High Cd toxicity but maintenance-free

- Nickel-zinc battery

$2NiOOH+2H_2O+2e \rightleftharpoons 2Ni(OH)_2+2OH^-$ $E^0=0.49$ V (cathode)

$Zn+2OH^- \rightleftharpoons ZnO+H_2O+2e$ $E^0=-1.24$ V (anode)

$2NiOOH+H_2O+Zn \rightleftharpoons 2Ni(OH)_2+ZnO$ $E_{cell}=1.73$ V (net)

Low cycle life but cost effective

Here M is a metal ion and z+ and (z+Δ)+ are its two oxidation states. For a material to qualify as a reversible electrode for secondary batteries it should satisfy the following conditions:

- As seen from reaction (I), the metal ion must be capable of adopting at least two oxidation states. This criterion is satisfied by many d-block elements.
- There should be a suitable chemical matrix that can host the metal ion in its multiple oxidation states. An oxide/hydroxide matrix is found to serve this purpose ideally although many sulphides are as good.
- The reaction (I) must have a high degree of reversibility which

- Nickel-metal hydride battery

$NiOOH+H_2O+e \rightleftharpoons Ni(OH)_2+OH^-$ E^0=0.49 V (cathode)

$MH+OH^- \rightleftharpoons M+H_2O+e$ E^0=–0.83 V (anode)

$NiOOH+MH \rightleftharpoons Ni(OH)_2+M$ E_{cell}=1.32 V (net)

High cycle life, high energy density, non-toxic, maintenance free

- Rechargeable alkaline manganese dioxide battery

$2MnO_2+H_2O+2e \rightleftharpoons Mn_2O_3+2OH^-$ E^0=0.26 V (cathode)

$Zn+2OH^- \rightleftharpoons ZnO+H_2O+2e$ E^0=–1.24 V (anode)

$Zn+2MnO_2 \rightleftharpoons ZnO+Mn_2O_3$ E_{cell}=1.5 V (net)

Shallow discharge but cost effective

- Lithium ion rechargeable battery

$LiCoO_2 \rightleftharpoons Li_{1-x}CoO_2+xLi^++xe$ E^0=0.6 V (cathode)

$Li_{1-x}C+xLi^++xe \rightleftharpoons LiC$ E^0=-3.0 V (anode)

$Li_{1-x}C+LiCoO_2 \rightleftharpoons LiC+Li_{1-x}CoO_2$ E_{cell}=3.6 V (net)

High energy density, high cycle life, maintenance free but uneconomical

- Zebra battery

$2Na^++NiCl_2+2e \rightleftharpoons 2NaCl+Ni$ E^0=-0.13 V (cathode)

$2Na \rightleftharpoons 2Na^++2e$ E^0=-2.71 V (anode)

$2Na+NiCl_2 \rightleftharpoons 2NaCl+Ni$ E_{cell} =2.58 V (net)

High temperature operation, still in development stages

The manner in which we are using secondary batteries today has not really solved the energy problem. The real breakthrough will come when advances in photovoltaics will enable us to recharge our secondary batteries by using solar energy.

is possible when neither the oxidised nor the reduced form is specially stable compared to the other. This is a difficult criterion to satisfy and only a few metals such as Pb, Mn, Ni and Co appear to pass this test.

- Both the reduced and oxidised forms of the material must have a fair degree of electrical conductivity or else the material will be rendered electrochemically inactive.
- Reaction (I) should have no competing reactions that produce electrochemically inactive side products which can reduce the activity of the electrode.

Several batteries have been in use commercially for a number of

Battery technologists must be ready with high performance, cost effective, environment friendly and socially acceptable batteries.

years. However, over the last two decades new applications and requirements have arisen which need to be met. As a result some old systems have declined in importance while others have expanded and new concepts have been developed. The commerically viable systems are: Lead-acid batteries, nickel-iron, nickel-cadmium, nickel-zinc, nickel-metal hydride, rechargeable alkaline manganese dioxide-zinc batteries, lithium ion rechargeable batteries and zebra batteries. The electrochemistry of these batteries is summarised in *Table 3*.

Future Prospects

Various battery technologies have been with us for many years. Technology dissemination and acceptance is no more a problem; in fact many new applications are crying out for the rapid development of new batteries with enhanced performance characteristics especially for communication, space, automotive, and traction purposes. Over the years many types of batteries have become commercially available. But the manner in which we are using secondary batteries today has not really solved the energy problem, as we recharge them using electrical energy obtained from fossil fuels. The real breakthrough will come when advances in photovoltaics will enable us to recharge our secondary batteries by using solar energy. By then battery technologists must be ready with high performance, cost effective, environment friendly and socially acceptable batteries.

Address for correspondence
A K Shukla
Solid State and Structural Chemistry Unit,
Indian Institute of Science,
Bangalore 560 012, India.

P Vishnu Kamath
Department of Chemistry,
Central College, Bangalore University,
Bangalore 560 001, India.

Suggested Reading

VS Bagotski. AM Skimdin. Chemical Power Sources (translated from the Russian by O Glebov and V Kisin). Academic Press. New York. 1980.

M Barak. Electrochemical Power Sources. Peter Peregrinus Ltd., Steverage. UK. 1980.

B D McNicol, D A J Rand (Ed.). Power Sources for Electric Vehicles. Elsevier. 1984.

D Linden (Ed.). Handbook for Batteries and Fuel Cells. McGraw Hill. New York. 1984.

***Journal of Power Sources*, 5 1:1,2. Elsevier. 1995.**

What's New in Computers

Video-on-Demand

M B Karthikeyan

This article presents an introduction to multimedia and describes one of its popular applications, video-on-demand.

M B Karthikeyan is working towards an M.Sc.(Engg.) at the Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore. His interests are multimedia storage systems, operating systems and networking.

What is Multimedia?

Traditionally, information services have been available to us in different ways. Educational material such as books, journals and newspapers are published and delivered mainly in printed form. Entertainment through television and radio broadcasts is produced and delivered primarily in analog form. Both these domains have undergone radical changes in recent years due to advances in computing and communication technologies. Innovations in digital signal processing, mass storage, and optical communication networks have enabled the integration of diverse types of media such as text, audio, video and graphics to be utilized in digital form. This integration is commonly referred to as *multimedia*. A system with the capability to capture, digitize, compress, store, retrieve, decompress and present multimedia information is called a *multimedia system*. *Figure 1* illustrates the main functional blocks of a simple multimedia system. Examples of multimedia systems include distance learning, home shopping, video-on-demand, video conferencing and information kiosks.

A system with the capability to capture, digitize, compress, store, retrieve, decompress and present multimedia information is called a *multimedia system.*

Characteristics of Multimedia

Media such as audio and video are called *continuous media*, because they consist of a sequence of media units (audio samples or video frames) that make sense only when presented in the same time sequence in which they were recorded. The design of information services to support continuous media differs significantly from services that allow only textual or numerical

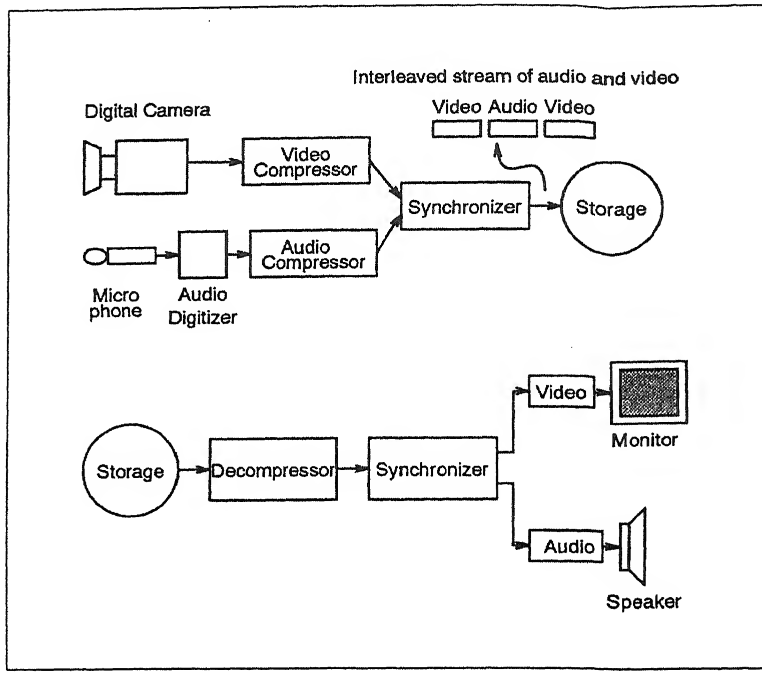


Figure 1 Capabilities of a multimedia system.

data, due to three fundamental characteristics of continuous media:

- *Storage and retrieval of continuous media are real-time operations.* Media capture devices (digital cameras and audio digitizers) generate a continuous stream of media units that must be stored in real time. During retrieval, the units of a continuous media stream are presented in the same time sequence in which they were captured. Any deviations from the timing sequence might result in perceptible glitches in audio and video presentations.

In spite of the stringent real time constraints and large storage and communication bandwidth requirements, multimedia applications are gaining popularity day by day.

- *Related media streams have to be temporally coordinated.* Several applications require the synchronization of more than one type of medium at the time of presentation. For example, a movie presentation requires the audio and video streams to be synchronized. This is also known as *lip synchronization*.

- *Continuous media have very high data transfer rates and large*

storage space requirements. Uncompressed digital audio and video have very high transfer rates and require large storage space. For example, the data rate for CD-ROM quality audio is 1.4 mega bits per second (Mbps; mega = 10^6). Several compression schemes have been developed to reduce the data size for the purposes of storage and transmission. MPEG (Motion Picture Experts Group) is an example of an international standard for digital video and audio compression.

In spite of the stringent real time constraints and large storage and communication bandwidth requirements, multimedia applications are gaining popularity day by day. To understand why, let us take a closer look at one popular multimedia application, namely *video-on-demand* or *interactive television*, that is likely to influence our daily lives in the near future.

What is Video-On-Demand?

The conventional mode of operation of television and cable television is the *broadcast* mode. In this mode, viewers have little flexibility in

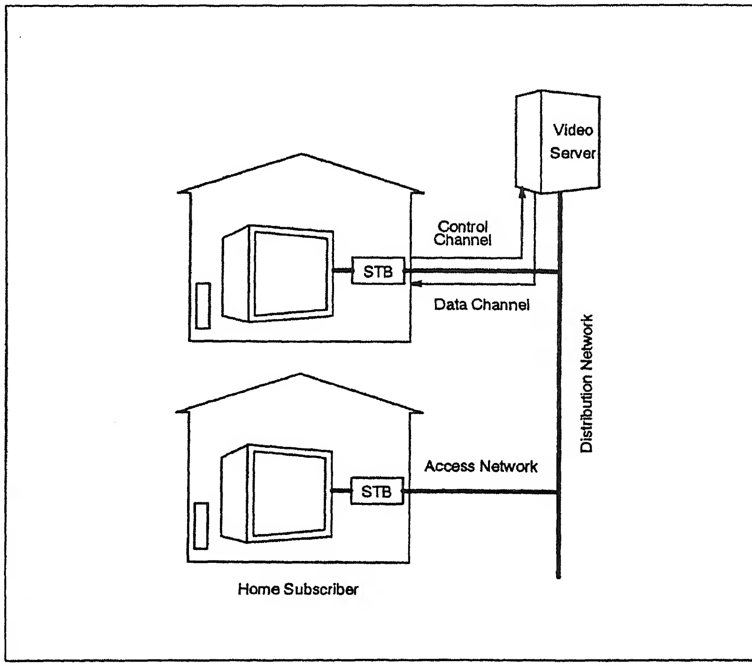
- Selecting programs (limited by the number of channels available),
- Scheduling the viewing time of programs, and
- Controlling the programs they view (viewers cannot skip parts of a program they find boring).

In contrast, an *on-demand* service provides its subscribers the ultimate flexibility in selecting programs when they wish, besides interactive control functions such as pause, resume, fast-forward, and rewind. The video rental service is in some sense an on-demand service. The electronic equivalent of the video rental service, that provides its subscribers on-demand access to a large collection of videos stored in high capacity servers over a broadband network is called a *video-on-demand* (VOD) system. Today's technology permits operators of telecommunication networks

Video Compression

The basic principle in audio and video compression is the elimination of redundant information. For example, in video, large portions of a scene remain unchanged from frame to frame. Video compression schemes exploit this fact by storing the differences (i.e., changes in positions of objects in a scene) between similar frames instead of the frames themselves. The more similar the frames the more compact the result. The maximum campression possible by MPEG-1 is about 200:1, but 50:1 is more typical. Newer compression methods have produced nearly 500:1 compression of video.

Figure 2 Components of a VOD system.



The electronic equivalent of the video rental service, that provides its subscribers on-demand access to a large collection of videos stored in high capacity servers over a broadband network is called a *video-on-demand* system.

(traditional telephone and cable TV operators) to provide a VOD service at a price competitive with the video rental service, without the need for the customer to travel.

Elements of a VOD System

Figure 2 illustrates the major elements of a VOD system. The three main components are: a video storage server, a network subsystem and customer premises equipment.

Video server: The video server consists of the storage and control required to store a large number of videos in compressed format and play back on request. It differs from traditional file servers in that it has to perform a number of functions such as

- *supporting continuous media storage and retrieval* – multimedia data is read or written by the server as a continuous stream of media blocks.
- *admission control* – the server checks if it has enough space and bandwidth to accommodate a new user session.

- *real-time request handling* – the server must ensure that each read or write request is completed within a time limit (request deadline).
- *guaranteed stream transmission* – to maintain a continuous display at the user's end, the server must transmit data at a steady rate.
- *stream encryption* – data streams transmitted over the network are scrambled or encrypted to prevent unauthorized access to the VOD service.
- *access control functions* – permits viewers to skip or replay portions of programs they watch.

A video server's functions are highly demanding in terms of storage space and transfer rate. For example, a two hour, MPEG-2 compressed movie (at an average transfer rate of 3 mega bits per second (Mbps)) requires 2.7 giga bytes (GB) (giga = 10^9) of storage. To store 500 such movies a video server would need 1.35 tera bytes (TB) (tera = 10^{12}) of storage. Because of these huge storage requirements, a video server is usually constructed as a hierarchy of storage media that includes semiconductor memory, secondary storage devices such as hard disks and tertiary storage devices such as optical jukeboxes. Of these the hard disks and optical jukeboxes form the bulk of the storage system. The easy availability of high performance hard disks permits the construction of large video servers based on an array of disks.

Space requirement

Assuming a data rate of 3 mega bits per second for TV quality video, the space required to store a one hour programme is $3 * 10^6 * 60 * 60 = 1.08 * 10^{10}$ bits; which is a lot of investment in terms of storage space.

In the normal mode of operation, admission control is done at the start of each customer session to determine if a new request can be serviced by the available resources in the system. If the new customer is admitted, media blocks of the requested video are retrieved from the storage system, temporarily buffered in memory and transmitted to the customer over the network. To guarantee uninterrupted playback the server must allocate storage bandwidth, buffer space and processing bandwidth.

Network subsystem: The network subsystem provides the interconnection of the various network elements in a VOD

system. Broadband networks enable video servers to communicate with the *set-top box* (STB) in the subscriber's home. These networks are *asymmetrical* in nature, with high bandwidth capacity from the video server to the set-top box (the downstream or data channel) and lower bandwidth for signalling from the set-top box to the video server (the upstream or control channel). Typically the bandwidth downstream is 1.5 Mbps (for one MPEG-1 stream) and upstream is 1.5 Kbps per user session.

Distribution networks carry video streams from the video servers to distribution points or head-end equipment by using a high speed transmission scheme such as *asynchronous transfer mode* (ATM) or synchronous optical network (SONET) which can provide output rates reaching 2.5 GBps. Access networks transport video from the head-end to set-top boxes by one of several alternative architectures: asymmetric digital subscribers loop (ADSL), hybrid fiber coax (HFC), or fiber in the loop (FITL). HFC is currently the popular architecture for the access network.

Adding interactivity to conventional broadcast television provides otherwise passive viewers the flexibility to choose programmes that match their interests and view them at times they wish.

The network's greater downstream bandwidth carries compressed audio and video streams, while the lower upstream channel carries the control signals from users to the video server. In addition, there are bi-directional control channels between the video server and the distribution and access networks to establish user sessions and reserve bandwidth for the downstream. The network ensures the real-time delivery of video streams at a constant rate (*isochronous* mode of operation), and the synchronization of related audio and video streams at the destination.

Customer Premises Equipment: The customer premises equipment consists of a set-top box (STB), a television monitor and a remote control. The STB is the bridge between the subscriber's display devices, peripherals, and input devices (such as a hand held infrared remote controller) and a communication channel in the access network. The channel connects the STB to

video or other information providers. The STB receives the incoming signal, demodulates it to recover the compressed digital video stream, decodes the stream and finally converts it to analog form and presents it to a TV monitor. The STB also has hardware blocks for remote control management, creation of necessary user interaction menus on the monitor and the sending of control information to the video server. Security and encryption are two other functions of an STB that prevent unauthorized access to services and ensure that subscribers are fairly charged. Future STB's might include a *personality module* that stores the subscriber's viewing profile, searches for and locates information that matches the subscriber's interests and schedules presentations at the subscriber's preferred viewing times.

Current technology trends indicate that by the turn of this century, what we see today as computer, television and telephone will all be wrapped into a single unit called a *teleputer*.

Basic Operation of a VOD Session

A typical VOD session consists of the following steps:

- *Connection establishment.* Initially, the set-top box is connected to a service gateway in the access network. From the gateway the user receives a list of accessible service operators.
- *Service selection.* The user selects a service operator from the list and the STB signals the selection to the gateway, which in turn selects the corresponding video server.
- *Path set-up.* Using the routing information in the service gateway, a path (downstream channel) is set up between the video server and the STB.
- *Program selection.* Once a video server is connected to the STB, a list of programmes offered by the server is downloaded to the STB. The browsing is done locally at the STB and requires no interactivity with the server. The user browses through the list and communicates his/her request to the video server via signalling.
- *User admission.* The video server on receiving the request determines if the new user can be serviced with the available resources. If sufficient bandwidth and storage capacity are avail-

able, the server starts transmitting the requested programme through appropriate actions in the network. Otherwise, the user request is queued up or cancelled.

- *Interactive program viewing.* The STB receives the programme from the network, decodes it and presents it to the viewer. Depending on the degree of interactivity supported by the video server and STB, the viewer can do some or all control functions such as pause, resume, fast-forward and rewind.
- *Session close.* The end of session is signalled by the STB to the video server, and resources allocated in the server and network are released.

VOD trial sites

One of Europe's largest trial systems began operation this January at Stuttgart, Germany and hopes to reach 4000 household and business subscribers. In the US, trial systems have been in operation for nearly a year now in Orlando, Omaha, Redmond, Fairfax County and San Jose. In England, VOD trials are underway in Colchester and Cambridge.

Conclusion

Digitization of diverse types of media such as text, audio, video and graphics has led to the development of a wide range of highly interactive multimedia services. The high level of interactivity is the key to the success of these services, in spite of the fact that multimedia data have unusually large storage space and communication bandwidth requirements. Adding interactivity to conventional broadcast television provides otherwise passive viewers the flexibility to choose programmes that match their interests and view them at times they wish. Present day computing and communication facilities allow interactive television services to be offered on a large scale and there are already several trial systems in operation in the US and Europe. Current technology trends indicate that by the turn of this century, what we see today as computer, television and telephone will all be wrapped into a single unit called a *teleputer*.

Address for correspondence
M B Karthikeyan
Supercomputer Education and Research Centre
Indian Institute of Science
Bangalore 560 012, India.

Suggested Reading

D Le Gall. MPEG: A Video Compression Standard for Multimedia. *Communications of the ACM*, 34(4). April 1991.

Winston William Hodge. Interactive Television - A Comprehensive Guide for Multimedia Technologists. McGraw-Hill. 1995.

Borko Furht, Deven Kalra, Fredrick L Kitson, Arturo A Rodriguez and William E Wall. Design Issues for Interactive Television Systems. *IEEE Multimedia*. May 1995.

In this section of Resonance, *we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. "Classroom" is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.*

! Energy transfer in an elastic collision

One may intuitively feel that in an elastic collision, energy is always transferred from an object of higher to one of lower energy. Interestingly, the physics and mathematics of collisions do not impose such a constraint. It is only necessary and sufficient that in such a process the total energy and momentum be conserved. We give here two examples of such processes to highlight this viewpoint.

In an elastic collision there need not be any energy transfer at all. As an example of this we mention the forward Compton scattering. Here an X-ray photon undergoes 'head on' collision with a static electron and emerges in the same direction without losing any energy to the electron.

Interestingly, in such a collision, energy can even be transferred from a body of lower to one of higher energy. Consider for example two mass points A and B with masses 0.5 gm and 1 gm and initial velocities 1cm/sec and 0.9 cm/sec respectively. Sooner or later A will overtake B and in the process collide with it. By the laws of conservation of energy and momentum after the collision, A acquires a velocity of $(13/15) \approx 0.8667$ cm/sec and

Is energy always transferred from an object of higher energy to one of lower energy in an elastic collision?

B a velocity of $(29/30) \approx 0.9667$ cms/sec. Clearly A has slowed down and B has speeded up. In this process thekinetic energy of B which was initially higher than that of A has actually increased after collision!

As a remark we may mention that if we average the values of an ensemble of collisions, then over all the more energetic particles lose energy and the less energetic ones gain energy. This is the statistical origin of equipartition (see *Am.J.Phys.* 62:487, June 1994).

Discussion of question raised in the Classroom section of **Resonance** *Vol.1, No.3.*

? An observer points a torch at a mirror that is moving away at velocity v. The frequency of the light is ν_0. The light reflected at normal incidence has a lower frequency ν. A calculation including special relativity, gives the result

$$\nu = \nu_0 \left(\frac{c - v}{c + v}\right)$$

where c is the speed of light. Why is this different from the standard Doppler shift formula for a source moving at a velocity v? This reads

$$\nu = \nu_0 \left(\frac{c - v}{c + v}\right)^{1/2}$$

Can one speak of a velocity of the image and if so what is it?

This question can be answered in two ways :

From Ritesh Kumar Singh, Ranchi; Mohan Devadass, Bangalore and Akshay Pundle (class XII), New Delhi.

a) From the velocity addition theorem of relativity we get the velocity of the image to be equal to $v' = (2v/(1 + v^2/c^2))$. Plugging this into the standard Doppler effect formula we get the required answer.

From Ritesh Kumar Singh, Ranchi

b) For an observer on the moving mirror the light (from the stationary observer) appears to have a frequency equal to $\nu' = \nu_0[(c - v)/(c + v)]^{1/2}$. Suppose this observer starts flashing light at this frequency ν'. Then the stationary observer will receive it as light of frequency $\nu'' = \nu'\ [(c-v)/(c+v)]^{1/2}$. Hence the result.

Think It Over

This section of Resonance *is meant to raise thought-provoking, interesting, or just plain brain-teasing questions every month, and discuss answers a few months later. Readers are welcome to send in suggestions for such questions, solutions to questions already posed, comments on the solutions discussed in the journal, etc. to* Resonance *Indian Academy of Sciences, Bangalore 560 080, with "Think It Over" written on the cover or card to help us sort the correspondence. Due to limitations of space, it may not be possible to use all the material received. However, the coordinators of this section (currently A Sitaram and R Nityananda) will try and select items which best illustrate various ideas and concepts, for inclusion in this section.*

1 Problem of the vacillating mathematician

From K B Athreya, Iowa State University

Assume that the home and office of a mathematician are one Unit apart. He starts from home in the morning for his office. Exactly halfway through he realizes that he has forgotten to bring something from home and starts to walk back towards home. Exactly halfway through (half the distance from the midpoint to home) he thinks that what was forgotten is not so important and starts to walk in the opposite direction. Halfway through this trip he again changes his mind and moves in the opposite direction. This continues. At each stage the distance covered is half of the distance from the point where he turns around to the home or the office depending on the direction of his present movement.

Where do you think this person will end up eventually? In other words, let x_n denote the distance from his home to the point where he is after his n th move. What are the limit points of this sequence?

Where will the vacillating mathematician end up eventually?

Questions posed by P Vishnu Kamath

2 Batteries

1 Why does one fill a lead-acid battery with distilled water?
2 Which of the batteries in Table 3 of the article on pages 66-67 of this issue is the most environment-friendly one?
3 What is the oxidation state of Ni in NiO(OH)?
4 How many different oxidation states can Mn exist in? Which of these are useful for the construction of a battery?
5 Can you name the reactions that occur at the battery electrodes when the battery is overcharged?
6 What will happen if you try to charge a primary cell?
7 From the answer to (6) above can you explain why certain cells carry a warning of possible explosions if you try to charge them?
8 Why can you not use a lead electrode in an alkaline battery and a Ni electrode in an acid battery?
9 Distinguish between a vented and a sealed battery; a flooded and a starved battery.
10 Calculate the total lighting load in your house, estimate your lighting energy requirement and calculate the weight of a lead acid battery to meet your requirements (assume the energy density of a lead-acid battery to be 40 Wh kg^{-1}).

Silver-zinc battery — a high power pack (courtesy ISRO)

How many times do I have to remind you that we are still in the stage of hunting — yet to be fit for those scientific queries!

The Origins of Science

Part II: After Thales

Gangan Prathap

In Part I of this essay, we had tried to locate a time, a place and a man in history from whom, one could argue, the great enterprise that we call science began. In this second and concluding part, we will examine the course of philosophy immediately after Thales' great intellectual leap.

Gangan Prathap is with National Aerospace Laboratories and Jawaharlal Nehru Centre for Advanced Scientific Research. His interests are: aircraft structural design and analysis, computational mechanics, philosophy of science and literature.

Introduction

If one has to look for a single agent who could be considered to be the founder of the philosophical and scientific tradition of the Western world, it would have to be Thales. So, we have a remarkable tradition of Greek philosophy, from Thales (580 BC) to Plato (430 BC), a tradition that has not been paralleled, let alone excelled, by any other period or any other culture or civilization. Sir Karl Popper, one of the leading philosophers of science of the 20th century, called this "the tradition of critical discussion". In all or almost all other civilizations, the scholastic tradition has been to pass on knowledge as doctrine, as received wisdom based on authority or text, from one generation to the next. The legacy of the Greeks, beginning with Thales and coming down through to Socrates, was that of persistent questioning, before the secrets of nature are yielded to us. Socrates said, "The unexamined life is not worth living."

The legacy of the Greeks, beginning with Thales and coming down through to Socrates, was that of persistent questioning, before the secrets of nature are yielded to us. Socrates said, "The unexamined life is not worth living."

From Thales to Anaximander

Thus, this newly emerging tradition encouraged criticism, even of one's masters. We find that the second of the great Ionian philosophers, Anaximander, also of Miletus and both pupil and kinsman of Thales, was ready with new ideas that were in direct conflict with the wisdom he received from his master. Popper

says that he likes to imagine that Thales was the first teacher to tell his pupils: "This is how I see things — how I believe that things are. Try to improve upon my teaching." On the other side of the world, around the same time, Buddha was to tell his disciples the same thing: "Don't accept anything as truth, even from your teacher, till you have verified it for yourself."

Anaximander speculated on the origin of the human race. He departed boldly from Greek mythology, which had so far traced the descent of man from the Gods and the titans. He made the heretical suggestion that all life emerged out of water and that man was once a fish. His biological speculations can be thought of as an early precursor of the theory of evolution.

Anaximander is remembered for having made the first map. He too addressed the question, "What is the world made of?" Unlike Thales, he argued that the ultimate physical reality could not be a physical substance itself. Thus, for water, he substituted an 'undefined something', an eternal imperishable substance with no properties but capable of containing 'oppositions' within itself such as hot and cold, wet and dry. He conceived of an eternal motion so that the familiar objects that appear to our senses are formed out of the 'oppositions' of this undefined substance and then return to it when they decay. Another idea he formulated was that of a balance of forces in nature. Thales believed that the flat earth rested on water, an argument which leads to an infinite regress. Anaximander was able to use his concept of the earth as being in the centre of an eddy or vortex which was in eternal motion so that it could remain freely suspended in space through a delicate balance of forces. This was a remarkable advance in pre-scientific thought. Anaximander also speculated on the origin of the human race. He departed boldly from Greek mythology, which had so far traced the descent of man from the Gods and the Titans. He made the heretical suggestion that all life emerged out of water and that man was once a fish.

We can immediately sense that Anaximander's ideas are a significant advance over Thales'. The element which Thales thought of as a first principle is shown to be derived from some indestructible primary matter. Here, Anaximander is willing to accept an abstract idea rather than settle for a more concrete substance like water. His biological speculations can be thought of as an early precursor of the theory of evolution.

Formulation of an Atomic Theory

Philosophers who followed Thales and Anaximander showed greater confidence and inclination to identify immortal principles and to justify these using systematic reason. To Anaximenes, who came immediately after Anaximander, the immortal principle was air. Notable among those who followed him are Parmenides and Zeno of the Eleatic school. A result of the severe logical examination of the physical theories that the Ionians produced was the formulation of the atomic theory. To us, some of the conclusions that they arrived at were strange (for example, that motion is an illusion; that empty space cannot exist and that the universe is a uniform distribution of matter!), but some of the intellectual processes they set in motion were to serve us well. One was a careful examination of the laws of logic. The other was the acceptance of the atomic theory, whereby everything in the universe is constituted from an infinite number of a single building block. Leucippus and Democritus are credited with this development and they also removed the earlier error of a motionless universe by arguing that natural motion of the constituent atoms is necessary to bring together and separate various forms of matter. There was therefore the possibility of cyclic change, of birth and decay, and all this reconciled with the idea of indestructible atomic constituents.

To Anaximenes, who came immediately after Anaximander, the immortal principle was air.

Change as the Essence of the Universe

The realization that reality was not something stable and change was the essence of the universe was made by Heraclitus. To Heraclitus, the immortal principle was fire. Heraclitus also suggested that all life 'evolved'; that everything grew in a constant state of flux. The world exists as a conflict and tension of opposites. This was to cause great consternation. When every thing was always changing, what could one say about anything? A distinction had to be made between the world of sense which was always changing, seemingly imperfect and unknowable with

Albert Einstein's View

"The development of Western science has been based on two great achievements, the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility of finding out causal relationships by systematic experiment (at the Renaissance). In my opinion one need not be astonished that the Chinese sages did not make these steps. The astonishing this is that these discoveries were made at all".

a presumed world of underlying reality which was unchanging, perfect and open to reason. This is the basis of the Platonic wisdom which was to rule western thought for nearly two thousand years. Again, what springs to mind is this disjunction — on the surface we have multiplicity and variety of appearances which are transitory but underneath, we reason that there is an inner and simplifying truth.

It is thus seen that the Greeks exalted reason above everything else. What we sense can be illusory but what we reason can converge to a picture of *nature* ruled by law, pattern and symmetry. It was the inner meaning that was more important than the outer appearance or event, and the Greeks believed that *a priori* reasoning could show us what the inner reality was. This was commendable but unscientific. Thus 'logos' or reason was often used and scientific exploration was excluded as something dispensable.

The Conflict Between Reason and Art

The conflict between what was 'logos' or reason, and what was 'techne' or art and craft was beginning to emerge. Roughly translated, technology is the practice of art and craft submitted to the scrutiny of reason. This does not make it science, yet. For science, as we shall see again and again, requires experimental proof.

Scientific hypotheses should be provisional theories formed to explain facts which are observed or observable, through practice or by experiment.

Kitto presents a very good illustration of this conflict by relating a controversy discussed by Hippocrates, who wrote the first great essay on medicine. Hippocrates was essentially protesting against the tendency of the *a priori* philosopher to disregard many facts which are obvious to one practising a craft, as medicine essentially was then. These men tended to frame general "hypotheses" using unsupported generalizations. This, Hippocrates argued, is not the way to deal with an art or craft where the methods and practices are well known and can point the way to many new discoveries. Scientific hypotheses should be provisional theories

formed to explain facts which are observed or observable, through practice or by experiment. The natural philosopher rejects and denies all this practical knowledge. Hippocrates was the first to point out the futility of such an attitude. In fact, it is this attitude that continues to give philosophy a bad name.

Lewis Wolpert's View

"Most of Greek science turned out to be wrong — for being wrong is a constant feature of the scientific method".

The Nature of Science

With Hippocrates, we see the first understanding of what the distinguishing feature of science is. In fact, this is to be the most important lesson we will learn from all our explorations into the philosophy of science. Thus, to paraphrase a definition from Kitto, there is science only where there is the possibility of building up a body of truth by observation and experiment; or to quote from a more original source, the Precepts of Hippocrates,

> "In medicine one must pay attention not to plausible theorizing ('logismos'), but to experience and reason ('logos') together. I agree that theorizing is to be approved, provided that it is based on facts, and systematically makes its deductions from what is observed. But conclusions drawn by the unaided reason can hardly be serviceable; only those drawn from observed fact."

Thus, there were Greeks who were truly scientific in the modern sense. Just as there must have been other equally bold pioneers in other cultures and civilizations. However, tragically, another tradition was to take firm root — philosophy began to supersede religion. Kitto discusses it extremely well. Greek philosophy's ambitious search for uniformity amidst the multiplicity of natural phenomena led to pure guess work (conjectures) and neglect of fact (refutation by experiment or observation in the attempt to frame comprehensive theories was totally dispensed with). Platonic idealism and the Aristotelian drive for comprehensiveness and dogmatism allowed generalizations to be made without empirical support or proof.

Platonic idealism and the Aristotelian drive for comprehensiveness and dogmatism allowed generalizations to be made without empirical support or proof.

While the Greeks 'shut their eyes', they kept their mind's eye

Science is a very fragile thing and there are ominous signs that it can fade away and a new dark age emerge to close our minds again.

open. A lot, in fact some say most, of Greek science turned out to be wrong. So what they failed to do in science, they more than made up for by their magnificent leaps into metaphysics, logic and mathematics. Euclidean mathematics and Archimedean mechanics are excellent examples of what they achieved. These were to prove to be fundamental to the future progress of science, but for this, one had to wait for another 1,800 years. Much of what Euclid and Archimedes laid down was forgotten and was preserved through the thoughtfulness of Greek and Islamic scholars of the Middle Ages. The availability of these writings to Galileo was to be another landmark in the history of modern science. Another important landmark is the use of the Socratic dialectical method, the search through logical enquiry, especially in matters of aesthetics and ethics. This is illustrated in the writings of Plato. Plato, says Kitto, "drew a sharp distinction between knowledge and opinion. Knowledge is not what a man has been told, shown or taught; it can be only what he has found out for himself by long and rigorous search." This could very well serve as the definition of what we mean by scientific research, giving up a life to intellectual striving so that the knowledge of 'what is' is yielded to us.

The Ionian school (or Milesian school of philosophy as it is sometimes called) was the first in which the pupils criticized their master. Anaximenes, who followed Anaximander, continued this critical tradition. Unfortunately, as Popper sadly notes, this critical and rationalist tradition was invented only once, and was to disappear with Plato and Socrates. The European mind was closed and remained closed until the critical tradition was rediscovered in Europe, during the Renaissance, nearly eighteen centuries later. That science appeared with the Ionians was a miracle. That it disappeared and remained dormant for eighteen hundred years was not surprising. That it was re-discovered, in a more vigorous aspect (reason and idealism was to join experience and empiricism) was another remarkable miracle for which we must be immensely thankful. However, we must be watchful as well - science is a very fragile thing and there are ominous signs that it can fade away and a new dark age emerge to close our minds again.

Suggested Reading

H D F Kitto. The Greeks. Pelican Books, Penguin. 1951.

Robert M Pirsig. Zen and the Art of Motor-Cycle Maintenance. Bodley Head. 1974.

Bryan Magee. Popper. Fontana Press. 1988.

John L Casti. Alternate Realities: Mathematical Models of Nature and Man. Wiley. 1989.

Lewis Wolpert. The Unnatural Nature of Science. Faber and Faber. 1992.

Jostein Gaarder. Sophie's World. Phoenix. 1995.

Address for correspondence
Gangan Prathap
Structures Division
National Aerospace Laboratories, PB No 1779
Bangalore 560 017, India.

Versatile Hydrogen

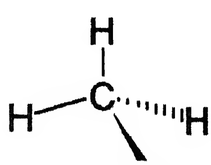
Hydrogen Bond with a Difference

A G Samuelson

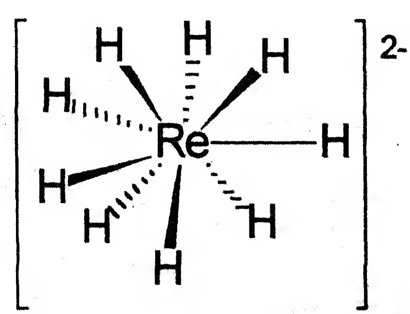
Hydrogen is probably the most intriguing element in the periodic table. Although it is only the seventh most abundant element on earth, it is *the* most abundant element in the universe. It combines with almost all the elements of the periodic table, except for a few transition elements, to form binary compounds of the type $E_n H_m$. Some of these compounds have fascinating structures (1, 2, 3).

However the most interesting interaction of hydrogen is the *hydrogen bond.* When a hydrogen atom is bound to an electronegative element it acquires a slight positive charge. As a result, it is attracted to other atoms such as nitrogen or oxygen in the vicinity, which have a slight negative charge. Therefore hydrogen bonds usually have the form X–H···Y in which X and Y are both electronegative atoms. A few examples of hydrogen bonds and their interaction energies are shown in *Table 1.* Although the strength of this interaction is relatively small in most cases, the presence or absence of hydrogen bonds has significant effects on the properties of molecules. Many interesting structural forms resulting from hydrogen bonds were discussed in the article by S Ranganathan (***Resonance***, Vol.1, 4, 1996).

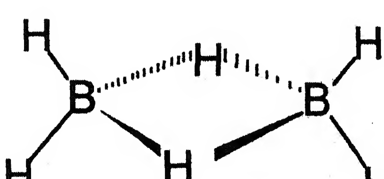
Very recently a new type of hydrogen bond has been experimentally characterised. In most metal hydrides, the hydrogen atom attached to the metal has a slight negative charge. It can be attracted to a hydrogen atom attached to nitrogen or oxygen. Hydrogen bond is formally formed in the unit X–H···H. A hydrogen attached to a metal acts as a hydrogen atom acceptor!



1



3



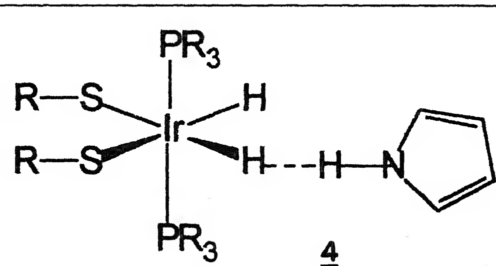
2

Table 1. Hydrogen Bonds and their strengths

H-Bond	Strength of the interaction in kJ mol^{-1}
$H_2S \cdots H\text{-}SH$	7
$H3N \cdots H\text{-}NH_2$	17
$H_2O \cdots H\text{-}OH$	22
$F\text{-}H \cdots F\text{-}H$	29

Over the last two years, several examples of such interactions have been documented. A typical case is the iridium hydride complex (**4**), hydrogen bonded to pyrrole.

An extremely simple and efficient method to estimate the strengths of this interaction has been evolved by Crabtree and Peris (E. Peris and R H Crabtree, *J C S Chem. Commun.*, 1995, 2179). They measured the energy associated with the stretching of the X–H bond using vibrational spectroscopy. When X–H interacts with an H-M unit, the X–H bond is weakened. The X–H stretching frequency is reduced. The position and also the width of this X–H vibrational band changes. By measuring the vibrational spectra for X–H in the presence and absence of M–H, it is possible to get reliable estimates of the strength of this X–H···H–M interaction. It is found to be a fairly weak bond with a strength of about 4-8 kJ/mol.



4

Metal hydrides and hydrogen bond donor molecules have been known for a long time. It is therefore surprising that the interaction between metal hydrides and X–H molecules has been established only recently. However, a plausible reason for this delay may be traced to a reaction of alkali metal and 1st row transition metal hydrides with H–X, where X is very electronegative. Since molecules containing X–H bonds are weak acids they react with the metal hydrides to give molecular hydrogen and M–X rather readily. In other words, the strength of the H···H interaction is so large that it ends up as H_2!

$$M\text{–}H + H\text{–}X \rightarrow M\text{–}X + H_2$$

So what is different in the recently discovered hydrogen bond? In these cases, the metal is from the 3rd or 4th row of the periodic table. The M–H bond is not too ionic and is sufficiently strong. So it engages in weak hydrogen bond

What is truly remarkable is the versatility of hydrogen in combining with electropositive and electronegative elements with equal ease, a fact that has often suggested a special place for it in the periodic table.

interactions without reacting to form H_2.

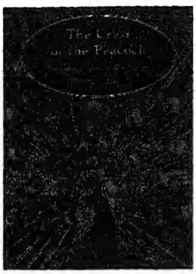
The same type of hydrogen bond has also been proposed for structures involving boron hydrides. The hydrogen atom of the B–H bond has an attractive interaction with hydrogen atoms attached to nitrogen or fluorine. For want of a better name, such X–H···H bonds have been recently christened *dihydrogen bonds.* Although at first it appears to be a new type of hydrogen bond, the principles involved in its formation are the same as in the classical hydrogen bond. The electronegative element Y carrying a partial negative charge in the usual H–bond has been replaced by a hydrogen that now carries the negative charge through its connection to an electropositive metal. What is truly remarkable is the versatility of hydrogen in combining with electropositive and electronegative elements with equal ease, a fact that has often suggested a special place for it in the periodic table.

A G Samuelson is with the Department of Inorganic and Physical Chemistry, Indian Institute of Science, Bangalore 560 012, India.

Non European Roots of Mathematics

Setting the Record Straight

R Sridharan



The Crest of the Peacock, Non-European Roots of Mathematics
George Gheverghese Joseph
Affiliated East-West Press,
62A, Ornes Road, Kilpauk,
Madras 600 010. 1990.
371pp. Rs. 150.

Giambattista Vico (1668-1744 A.D.), the Neapolitan philosopher of history, in his *Scienza Nuova*, cautions against two pitfalls which the scholars should guard against while writing history. The first is what he calls 'conceit of nations' which makes one adopt one's national point of view while writing it. The second is 'conceit of scholars', which makes one interpret historical evidence on the assumption that the criteria of rationality and correctness according to one's own present condition should equally apply to all the previous ages. Unfortunately, these two characteristics generally dominated the Western analysis of the contributions of the non-Western societies to science in general and mathematics in particular.

One could perhaps dismiss as an example of a colonialist's attitude, G.R. Kaye's insistence during the early part of the twentieth century,

Reprinted from *Current Science*

> The aim of this book is to put up an aggressive defence against 'Eurocentrism', to underscore the 'non-European mathematical achievements' and to 'dent the parochialism that lies behind the eurocentric perception of the development of mathematical knowledge'.

that the Bakshali manuscript was no older than the 12th century A.D. (against all accepted evidence). How does one explain the attitude of Van der Waerden (*Russian Mathematical Surveys*, 1976, **31**) who claimed as recently as a few years ago, that the idea behind the ingenious Chakravala method of the Indians was most probably borrowed from the Greeks, except as wishful thinking?

The aim of George Gheverghese Joseph in *The Crest of the Peacock* is to put up an aggressive defence against 'Eurocentrism', to underscore the 'non-European mathematical achievements' and to 'dent the parochialism that lies behind the eurocentric perception of the development of mathematical knowledge'.

Here is a summary in brief of the contents of this book. The Ishango bone with its markings, discovered in central equatorial Africa, is a riddle; the markings possibly representing a lunar calendar or a calendar of events of a ritual nature recorded by humans more than several thousand years ago. The *Quipu* of the Inca Indians (which consists of a collection of coloured cords with knots), represents a very

sophisticated device for storing numerical information. The first credit for written mathematics including the written number systems (in the hieroglyphic, hieratic and demotic notation) goes to Egypt. (The Aztecs of central America also had a system of numerals similar to the Egyptians.) During the long period of ancient Egyptian history, Egypt made several contributions to mathematics, particularly to the representation of fractions and to calculations of areas and volumes. The history of ancient Mesopotamia, dating back to 3500 B.C., had a long chain of various ruling dynasties. The most outstanding achievements of the Babylonian civilization were in various numerical and algebraic problems. They were certainly aware of the Pythagorean triples and the notion of similarity of triangles.

As in many other things, mathematics in China has had a very long history and it developed mostly in isolation. The Chinese had a system of notation for using nine numerals as early as the 14th century B.C. and the theorem of Pythagoras was well known to them from ancient times. They studied indeterminate equations of the first degree in their construction of calendars. The so called *Horner's method* for solving polynomial equations was already known to the Chinese very much earlier.

The first credit for written mathematics including the written number systems goes to Egypt. During the long period of ancient Egyptian history, Egypt made several contributions to mathematics, particularly to the representation of fractions and to calculations of areas and volumes.

The author begins his survey of the Indian contributions to mathematics with the Harappan civilization and its advanced brick-making technology. The building of sacrificial altars during the Vedic period led to the geometry of the Sulva Sutras which show knowledge of irrational numbers, geometrical figures, Pythagoras theorem, etc. The heretical sects like the Jains contributed equally to the growth of mathematics. For instance they considered laws of indices, problems in permutations and combinations, progressions, etc. The work of Pingala on metres contains the beginnings of binary arithmetic. The author also includes a section on the construction of Sriyantras (associated with tantric practices) and their geometry. Somewhat curiously, he also discusses the book *Vedic Mathematics* by Bharati Krishna Tirthaji and in particular the 'Nikhilam method' for multiplication which Tirthaji attributed to the Vedas.

The Bakshali manuscript which is probably a later version of a document dating back to the first few centuries of the Christian era bridges the gap between the mathematics of the Sulva period and that of the classical period.

Mathematics of the classical age began at Kusumapura (modern Patna), where, for

> It was once upon a time believed that mathematics in India essentially came to a halt after Bhaskara. But as has been ably shown by modern research, mathematical tradition continued through the middle ages in Kerala, till in fact the 16th century.

instance, the great Jain metaphysician, Umasvati, worked during the second century B.C. Here Aryabhata (c 476 A.D.) wrote his classical work, *Aryabhatia*, on astronomy (which contains incidentally the Kuttaka method of solving linear indeterminate equations). The classical age can boast of a series of successors of Aryabhata: Brahmagupta (b 598 A.D.) was not only a great astronomer but also a remarkable mathematician, who made fundamental contributions to the study of quadratic indeterminate equations, composition of binary quadratic forms, geometry of cyclic quadrilaterals, etc., Sridhara (c 900 A.D.), author of standard works on arithmetic and mensuration; and finally Bhaskara II (b 1114 A.D.) of the Ujjain school, who is very well known as the author of *Lilavati* and *Bijaganita* (parts of *Siddhanta Shiromani*). It was once upon a time believed that mathematics in India essentially came to a halt after Bhaskara. But as has been ably shown by modern research, mathematical tradition continued through the middle ages in Kerala, till in fact the 16th century. Narayana Pandita, Madhava and Neelakanta Somayaji are some of the well-known mathematicians of this period, who worked extensively on trigonometric series and calculus. The *Yukti Bhasha* of Jyeshta Deva (c 1550 A.D.) is a notable summary (in Malayalam) of the mathematical achievements of the Kerala School.

One should also mention the name of an otherwise unknown mathematician Jayadeva (c 1000 A.D.), who is mentioned in a commentary, by Udayadivakara, entitled *Sundari of Laghu Bhaskariya* (by Bhaskara I, 629 A.D.) for his complete solution of quadratic indeterminate equations by the remarkable Chakravala method.

The book ends with a detailed account of the contributions of the Arabs to mathematics. The author takes great pains to show that the generally-held belief that their contributions consisted merely in transmitting mathematical knowledge is not correct. In fact they made a happy fusion of the geometric spirit of the Greeks with the algorithmic traditions of Babylonia, India and China, making substantial contributions to mathematics.

I would like to end with a few comments on

> The generally-held belief that Arab contributions consisted merely in transmitting mathematical knowledge is not correct. In fact they made a happy fusion of the geometric spirit of the Greeks with the algorithmic traditions of Babylonia, India and China.

It is true that all the civilizations of the past have thought about questions in arithmetic. But it cannot be denied that modern mathematics, as it is understood today, does owe a great deal to the renaissance in Europe, which in turn was a miraculous revival of Greek thought.

this thought-provoking and informative book. It is true that all the civilizations of the past have thought about questions in arithmetic. But it cannot be denied that modern mathematics, as it is understood today, does owe a great deal to the renaissance in Europe, which in turn was a miraculous revival of Greek thought. The Greeks held mathematics (particularly geometry) on a high pedestal and regarded it as a noble human endeavour. It is therefore not surprising (as is recorded by Plutarch in his *Life of Marcellus*) that Plato was critical of Eudoxus and Archytas for their use of mechanics for proving geometrical theorems. Quite apart from being a tool in technology, Mathematics has an innate beauty of its own which gives it a nobility, permanence and remoteness, which I do not see why the author should resent.

The author sometimes overstates his case. For instance, he includes the book *Vedic Mathematics* in his discussion of the mathematics of the Vedic period while there is so far no basis at all to Tirthaji's claim that its mathematics goes back to the Vedas. While discussing the Chakravala method (perhaps the most beautiful contribution of ancient India to modern mathematics), the author refers to Selenius' paper (1975) and does not mention that the commentary *Sundari* was discovered by Kripa Shankar Shukla (c.f. *Ganita* , 1954, p.1-20) and that he was the first to highlight the contribution of Jayadeva, I hope that at least in the next edition, due credit is given to Shukla for this discovery.

The above are only minor criticisms of this eminently readable and knowledgeable book, which is born out of a labour of love. While one may not agree with all the conclusions arrived at by the author, one has to accept that he has made an excellent case for the existence of non-European roots for mathematics. He deserves praise for his admirable effort.

R Sridharan is with Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India.

Supercool theory solves hot ice cream puzzle ... How is it possible for hot water to freeze more quickly than cold? This peculiar phenomenon, first noticed by Aristotle in the 4th century BC, has baffled scientists for generations. But a South African physicist now claims that the answer lies in water's ability to remain liquid below its normal freezing point. "It's all to do with supercooling," says David Auerbach, who works at the Max Planck Institute for Fluid Dynamics in Germany. *(New Scientist, 2 December 1995).*

Guidelines for Authors

Resonance - *journal of science education* is primarily targeted to undergraduate students and teachers. The journal invites contributions in various branches of science and emphasizes a lucid style that will attract readers from diverse backgrounds. A helpful general rule is that at least the first one third of the article should be readily understood by a general audience.

Articles on topics in the undergraduate curriculum, especially those which students often consider difficult to understand, new classroom experiments, emerging techniques and ideas and innovative procedures for teaching specific concepts are particularly welcome. The submitted contributions should not have appeared elsewhere.

Manuscripts should be submitted in *duplicate* to any of the editors. Authors having access to a PC are encouraged to submit an ASCII version on a floppy diskette. If necessary the editors may edit the manuscript substantially in order to maintain uniformity of presentation and to enhance readability. Illustrations and other material if reproduced, must be properly credited; it is the author's responsibility to obtain permission of reproduction (copies of letters of permission should be sent). In case of difficulty, please contact the editors.

Title Authors are encouraged to provide a 4-7 word title and a 4-10 word sub-title. One of these should be a precise technical description of the contents of the article, while the other must attract the general readers' attention.

Author(s) The author's name and mailing address should be provided. A photograph and a brief (in less than 100 words) biographical sketch may be added. Inclusion of phone and fax numbers and e-mail address would help in expediting the processing of manuscripts.

Summary and Brief Provide a 2 to 4 sentence summary, and preferably a one sentence brief for the contents page.

Style and Contents Use simple English. Keep the sentences short. Break up the text into logical units, with readily understandable headings for each. Do not use multiple sub sections. Articles should generally be 1000-2000 words long.

Illustrations Use figures, charts and schemes liberally. A few colour illustrations may be useful. Try to use good quality computer generated images, with neatly labelled axes, clear labels, fonts and shades. Figure captions must be written with care and in some detail. Key features of the illustration may be pointed out in the caption.

Boxes Highlights, summaries, biographical and historical notes and margin notes presented at a level different from the main body of the text and which nevertheless enhance the interest of the main theme can be placed as boxed items. These would be printed in a different typeface. Such a boxed item should fit in a printed page and not exceed 250 words.

Suggested Reading Avoid technical references. If some citations are necessary, mention these as part of the text. A list of suggested readings may be included at the end.

Layout It is preferable to place all the boxes, illustrations and their captions after the main text of the article. The suggested location of the boxes and figures in the printed version may be marked in the text. In the printed version, the main text will occupy two-thirds of each page. The remaining large margin space will be used to highlight the contents of key paragraphs, for figure captions, or perhaps even for small figures. The space is to be used imaginatively to draw attention to the article. Although the editors will attempt to prepare these entries, authors are encouraged to make suitable suggestions and provide them as an annexure.

Book Reviews

The following types of books will be reviewed : (1) text books in subjects of interest to the journal; (2) general books in science brought to the attention of students/teachers; (3) well-known classics; (4) books on educational methods. Books reviewed should generally be affordable to students/teachers (price range Rs.50 to 300).

New books will get preference in review. A list of books received by the academy office will be circulated among the editors who will then decide which ones are to be listed and which to be reviewed.

Acknowledgements

Resonance gratefully acknowledges the help received from the following individuals:

V Pati, Jayant Rao, V Ravi, Joseph Samuel and R Vasundhara.

We thank *Current Science* for permission to reproduce R Sridharan's review of the book *The Crest of the Peacock* and Physical Research Laboratory, Ahmedabad for sending us photographs of the Hyakutake comet taken at Gurushikhar Observatory, Mount Abu.

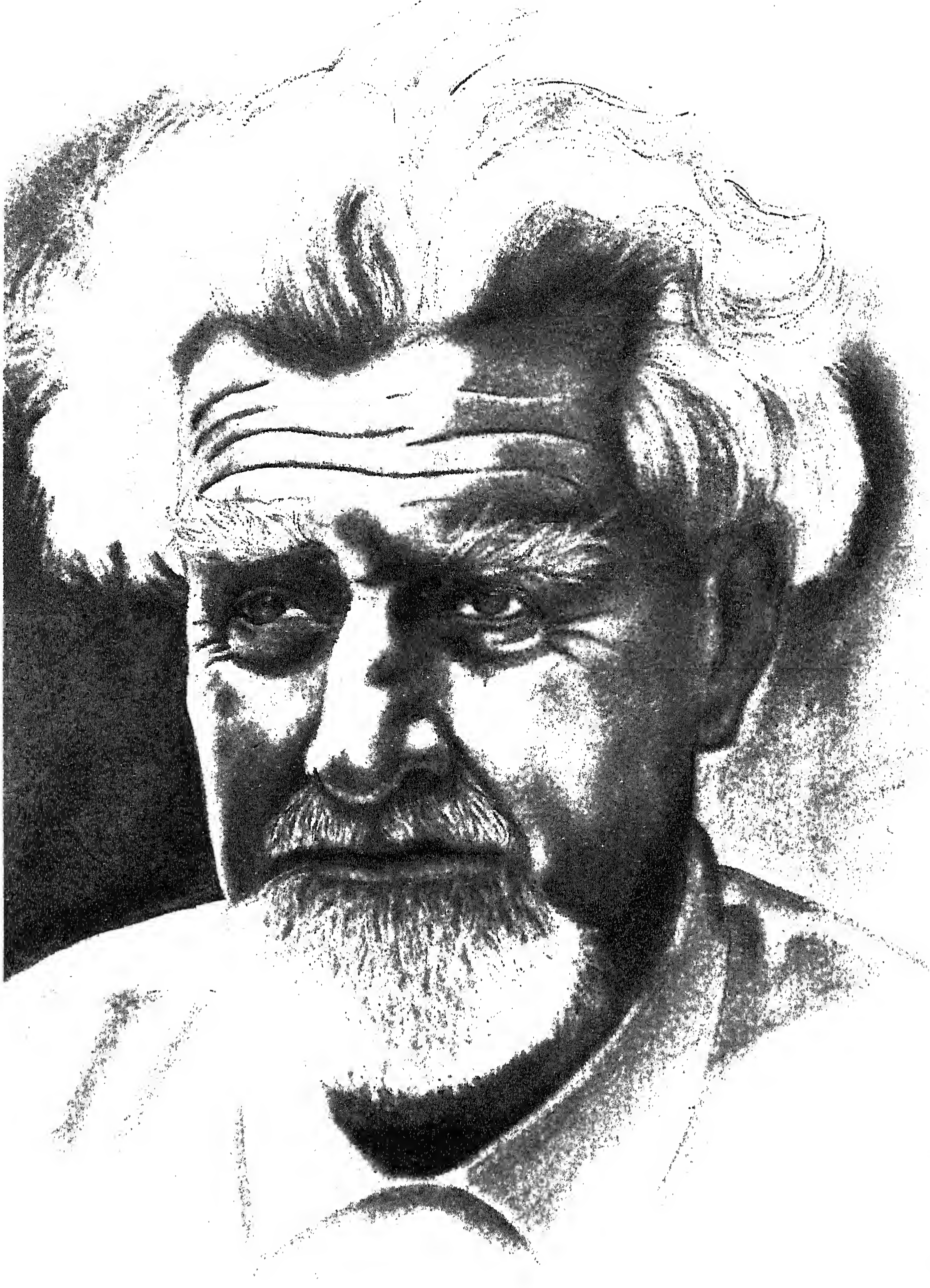
Annual Subscription Rates

	Personal	Institutional
India	Rs 100	Rs 200
Third World Countries	US $ 25	US $ 50
Other Countries	US $ 50 (Students $ 25)	US $ 100

Send your subscriptions by DD or MO in favour of "Indian Academy of Sciences" to Circular Department, Indian Academy of Sciences, C V Raman Avenue, PB No. 8005, Bangalore 560 080, India.

Edited and published by V K Gaur for the Indian Academy of Sciences, Bangalore 560 080. Printed at Thomson Press (I) Ltd., Faridabad 121 007.

Konrad Zacharia Lorenz (b. 7 November 1903, d. 27 February 1986) along with Niko Tinbergen and Karl von Frisch laid the foundations for the scientific study of animal behaviour or ethology and shared the Nobel Prize with them for doing so.



Konrad Zacharia Lorenz
1903-1986